

AD_____

Award Number: DAMD17-01-1-0724

TITLE: Telomerase Independent Telomere Maintenance in Ovarian
Cancer: A Molecular Genetic Analysis

PRINCIPAL INVESTIGATOR: Dominique Broccoli, Ph.D.

CONTRACTING ORGANIZATION: Fox Chase Cancer Center
Philadelphia, Pennsylvania 19111

REPORT DATE: July 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20041214 074

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503				
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE July 2004	3. REPORT TYPE AND DATES COVERED Final (1 Jul 2001 - 30 Jun 2004)		
4. TITLE AND SUBTITLE Telomerase Independent Telomere Maintenance in Ovarian Cancer: A Molecular Genetic Analysis		5. FUNDING NUMBERS DAMD17-01-1-0724		
6. AUTHOR(S) Dominique Broccoli, Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Fox Chase Cancer Center Philadelphia, Pennsylvania 19111 E-Mail: Dominiuge.broccoli@fccc.edu		8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		10. SPONSORING / MONITORING AGENCY REPORT NUMBER		
11. SUPPLEMENTARY NOTES Original contains color plates: All DTIC reproductions will be in black and white.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 Words) The goal of this project is to elucidate some of the genetic and biological determinants of ovarian cancer, focusing on an in vitro model for ovarian cancer that we have developed. We found that immortalization and transformation of human ovarian surface epithelial (HOSE) cells can differ in the pathway used for telomere length maintenance, a phenomenon that we have also observed in the clinical disease. We have found that the majority of our HOSE cell cultures use the Alternative Lengthening of Telomeres (ALT) pathway for telomere maintenance, thereby providing an in vitro model to characterize the underlying basis of the ALT pathway in ovarian cancer. We completed Task 1 having characterized telomere dynamics in 50 ovarian tumors, created a number of cDNA custom microarrays and used these resources to assess expression differences in clinical samples. As proposed in Task 2, we successfully reintroduced telomerase into several HOSE cell lines and have identified two telomeric factors, TRF2 and Tankyrase2, whose expression is altered following forced expression of telomerase in ALT positive cells. Finally, as intended in Task 3 we have demonstrated that the tumor suppressor p53 negatively regulates ALT, the first identification of regulatory requirements for ALT.				
14. SUBJECT TERMS Alternative lengthening of telomeres (ALT), telomerase independent telomere maintenance, cDNA microarray analysis of differentially expressed genes, human ovarian surface epithelial (HOSE) cells, ovarian cancer. p53			15. NUMBER OF PAGES 42	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

Table of Contents

Front Cover	1
Standard Form 298	2
Table of Contents	3
Introduction	4
Body	4
Key Research Accomplishments	7
Reportable Outcomes	8
Bibliography of Publications	8
Conclusions	9
References	9
List of Personnel	11
Appendices	11

Introduction

The diseases that are commonly referred to as ovarian cancer in the vast majority of cases develop from the malignant transformation of a single cell type, the surface epithelium. However, the biological mechanisms leading to transformation remain unclear. The goal of this project is to elucidate some of the genetic and biological determinants of ovarian cancer, focusing on an *in vitro* model for ovarian cancer that we have developed (1-3). We have initiated primary human ovarian surface epithelial (HOSE) cell cultures and have successfully derived HOSE cell lines that have undergone immortalization and spontaneous transformation *in vitro* and can form tumors *in vivo*. Furthermore, we found that immortalization and transformation of HOSE cells can differ in the pathway used for telomere length maintenance (4), a phenomenon that we have also observed in the clinical disease. Maintenance of telomeric repeats is required for immortalization and is commonly associated with activity of telomerase [reviewed in (5)]. However, a number of tumors and tumor cell lines have been described that do not have telomerase activity in which the telomere length dependent limitation on cell division is circumvented by a mechanism called Alternative Lengthening of Telomeres (ALT) (6, 7). We have recently found that 30% of advanced stage ovarian adenocarcinomas lack telomerase activity and thus would be refractory to treatment with telomerase inhibitors. The ALT pathway also represents a salvage pathway that may be activated in tumors in response to short telomeres arising as a consequence of telomerase inhibition. We have found that the majority of our HOSE cell cultures use the ALT pathway for telomere maintenance (4), thereby providing an *in vitro* model to characterize the underlying basis of the ALT pathway in ovarian cancer. The mechanism(s) leading to ALT is unknown yet is clearly important in tumorigenesis. Using our HOSE tissue culture model and powerful molecular genetic screening tools we propose to uncover genes of relevance to ALT and malignant transformation of the ovarian surface epithelial cells.

Body

Task 1: Identifying genes that are differentially expressed in ALT HOSE cells and determining the expression pattern of candidate ALT-causing genes in a panel of immortal cell lines and ovarian tumor samples that lack telomerase expression. We have completed the research initially proposed in this aim. Progress made here has led to ongoing research efforts to continue to identify factors essential for telomere maintenance by ALT.

In this study, eight SSH libraries were derived. Specifically, RNA was pooled from three telomerase negative/ALT positive HOSE cells (*i.e.*, HIO-107, 118, and 135) and from two telomerase positive/ALT negative HOSE cells (*i.e.*, HIO-80 and 114). The first library was generated to identify genes that should be overexpressed in the ALT-positive versus ALT-negative cells. The second library was generated to uncover genes that should be overexpressed in ALT-negative versus ALT-positive cells. Libraries 3 and 4 were generated by pooling RNA from telomerase negative/ALT positive HOSE cells (*i.e.*, HIO-107, 118, and 135) and from their mortal counterparts (telomerase negative/ALT negative). Libraries 5 and six were generated by pooling RNA from telomerase positive/ALT negative HOSE cells (*i.e.*, HIO-80 and 114) and from their mortal counterparts (telomerase negative). Finally, libraries 7 and 8 were generated by pooling RNA from two HIO-118 clonal cell lines in which hTERT had been reintroduced and from their parental HIO-118 line. The cDNAs generated from each subtraction were subcloned into plasmids and the plasmids were then transfected into bacteria. Bacterial clones were selected and arrayed into 96 well microtiter plates. The bacterial arrays were duplicated and the cDNA inserts were amplified by polymerase chain reaction (PCR). The PCR fragments were evaluated by gel electrophoresis and the DNA purified and arrayed into 96 well microtiter plates. The cDNA clones were then doubly spotted onto glass slides. RNA from mortal and immortalized HOSE cells was hybridized to our custom cDNA microarray slides.

To expand our screening possibilities, we have also created glass array containing 10,000 "named" genes. We have also interrogated glass slides containing 40,000-element cDNA microarrays.

For these experiments we utilized clones from a Research Genetics sequence verified cDNA library. RNA samples are being used in parallel to evaluate both our custom cDNA and "named" cDNA microarrays. We have isolated several dozen genes with distinctive and consistent patterns of expression (*e.g.*, either lost or gained upon immortalization, expressed in ALT versus telomerase positive). These genes are currently being evaluated in clinical samples by Northern blotting and RT-PCR approaches.

We have completed characterization of an additional 50 ovarian tumors for telomerase activity versus ALT. Of the 50 tumors characterized, 28 contained detectable telomerase activity using 0.5 μ g of extract. The remaining tumor extracts did not exhibit telomerase activity. To confirm whether these tumors were in fact telomerase negative, additional TRAP assays using a range of extract (from 0.05 μ g to 5 μ g) were carried out. In addition, telomerase negative tumors were reassessed following preparation of a second extract. Upon completion of this analysis an additional 9 tumors were determined to contain telomerase activity. Therefore, 36 out of 50 ovarian tumors analyzed (72%) were found to be telomerase positive. Southern blot analysis was carried out to determine telomere length in 9 of the 14 telomerase negative tumors where sufficient DNA was available. In addition, telomere length was determined for 16 of the telomerase positive tumors. This analysis yielded two tumors with the ultra-long telomeres (>20Kb) that are characteristic of the ALT pathway (6). The remaining 7 telomerase negative tumors had telomere lengths similar to those seen in telomerase positive tumors.

Cell lines that utilize the ALT pathway for telomere maintenance contain large multiprotein complexes in which telomeric proteins and DNA co-localize with the PML nuclear body, called ALT-associated PML nuclear bodies (APBs) (8). To determine if the tumors with telomere lengths within the size range typically seen of telomerase positive tumors were also using the ALT pathway for telomere maintenance, we are analyzing these tumors for the presence of APBs. This analysis required optimization of the ability to detect these structures in paraffin embedded tissue sections. This technical optimization has been carried out and we are able to detect APBs in paraffin embedded sections. However, we have not detected APBs in the limited subset of ovarian tumors tested. ALT has recently been reported to be activated frequently in sarcomas (9, 10) and to that end we have begun an analysis of ALT in these tumors using the technologies developed through DOD funding. Rapid progress has been made and we can definitively demonstrate ALT in approximately 20% of the liposarcomas tested to date.

Finally, we have improved our ability to profile the RNA expression patterns of tumors. First, we have derived oligonucleotides microarray. We have found that oligonucleotides microarrays give superior quality of hybridization signal relative to cDNA arrays by reducing non-specific binding and thus general background. These arrays contain 29,952 oligonucleotides that have been spotted onto polylysine coated microscopic glass slides. Each oligo in the set is 50 bp in length, correspond to HPSF® (High Purity Salt Free) quality standard and was designed to a coding region of a unique gene. Arrays have been spotted using a GeneMachine Omnigrid arrayer (GeneMachine, San Carlos, CA). From each batch of slides, one undergoes strict quality control for spot morphology and DNA abundance. Secondly, we have refined methods to amplify the RNA from microdissected tumor cells. We have recently evaluated RNA from ~10,000 microdissected cells.

We will extend our analysis of factors involved in activation and maintenance of the ALT pathway using the sarcoma samples and the techniques developed under the auspices of DOD funding. Importantly, it has recently been demonstrated that ALT is activated in cell culture model systems following telomerase inhibition. These data indicate that understanding the basis of ALT is of critical importance as telomerase inhibitors are currently being tested in clinical trials.

Task 2: To determine the telomere dynamics in ALT positive HOSE cells following exogenous expression of hTERT and identify genes that are differentially expressed in ALT HOSE cells following expression of hTERT. Over the course of funding from the DOD we have completed this task entirely

and have identified two telomeric factors whose expression is altered following forced expression of telomerase in ALT positive cells.

The catalytic subunit of telomerase, hTERT, was introduced into two cell lines derived from human ovarian surface epithelium (HOSE), HIO118 and HIO107, by transfection to induce telomerase activity. These cell lines have previously been characterized and demonstrated to use the ALT pathway for telomere maintenance (4). Three cell lines, HIO118A5, HIO118D6 and HIO107C6, were identified that contained telomerase activity. Two cell lines, HIO118A6 and HIO107B5, that were hygromycin resistant but did not exhibit detectable telomerase activity were also isolated. These lines provided internal controls for alterations observed in the telomerase positive clones isolated from the same transfection. The level of telomerase activity present in the HIO118A5, HIO118D6 and HIO107C6 cell lines was quantitated relative to the levels of telomerase activity in HeLa cells. The HIO118A5 and HIO107C6 cell lines had relatively stable levels of telomerase activity over the time course of these experiments. For the HIO118A5 cell line telomerase activity ranged from 1.2 to 1.5 times that present in HeLa cell extract, while for the HIO107C6 cell line telomerase activity ranged from 0.5 to 0.8 times that present in HeLa cell extract. In contrast, the HIO118D6 cell line initially had almost 1.8 times the level of telomerase activity as is present in HeLa cells; this activity gradually decreased over time in culture until it was completely absent by PD80.

Telomere length in each clonal line was analyzed by Southern blot analysis at various times during culture. At the earliest possible point for analysis, PD 16, telomere length in both HIO118 telomerase positive subclones, HIO118A5 and HIO118D6, had altered from that in the parental HIO118 cell line exhibiting an increased hybridization intensity of lower molecular weight DNA fragments. These data indicate that the presence of telomerase activity alters the telomere size distribution in ALT cells.

Cell lines that utilize the ALT pathway for telomere maintenance contain large multiprotein complexes in which telomeric proteins and DNA co-localize with the PML nuclear body, called ALT-associated PML nuclear bodies (APBs) (8). To determine if forced expression of telomerase in ALT cell lines resulted in inhibition of this marker of the ALT pathway, we carried out indirect immunofluorescence and determined the frequency of these structures in logarithmically growing cells. All of the clonal cell lines generated here contained APBs, irrespective of whether or not they had detectable telomerase activity.

Telomeres serve the essential function of providing stability to ends of linear chromosomes. One phenotype exhibited by cells that have lost telomere end-protection function is the end-to-end fusion of chromosomes (11). During anaphase, these fusions are manifested as bridges of unresolved DNA between the separating daughter nuclei. To determine if expression of telomerase resulted in increased telomere stability, we analyzed the frequency of anaphase cells that exhibited fusions in the parental HIO118 and HIO107 cell lines, all hTERT derivative cell lines, and an independently derived cell line, HIO114, that activated telomerase spontaneously upon immortalization. The parental ALT cell lines have a higher frequency of anaphase bridges than does HIO114, consistent with there being a higher level of telomere malfunction in these cell lines. The frequency of anaphase bridges in the clonal lines analyzed here was similar regardless of the presence of telomerase activity. The lowest frequency of bridges (reduced by almost 2 fold from that in the parental cells) was present in the HIO118A5 cell line, which maintained the highest consistent level of telomerase activity during the course of these experiments. However, the HIO118A5 cell line still had a higher frequency of anaphase bridges than did the telomerase positive HIO114 cell line. In the HIO118D6 cell line anaphase bridges occurred at a frequency similar to that in the parental cell line, although this cell line also had relatively high levels of telomerase activity (0.8 times of that present in HeLa cells at the time of analysis). The HIO107C6 cell line did not exhibit an alteration in the frequency of anaphase bridges relative to the HIO107 parental cell line or to the telomerase negative HIO107B5 cell line.

To identify genes potentially involved in ALT- or telomerase-dependent mechanisms of telomere maintenance, we compared gene expression profiles of parental ALT and derived telomerase-expressing HIO cell lines. Using 40,000-element cDNA microarrays we compared gene expression profiles of HIO118 A6 parental ALT versus HIO118 A5 and HIO118 D6 derived *hTERT* expressing cell lines. Within the subset of 250 genes (~0.6%) with expression level differences ranging from 3 to 67-fold, were those encoding hTRF2 and tankyrase 2, both telomeric-binding proteins involved in telomere protection and maintenance (11-14). Both genes were upregulated in telomerase-reconstituted cell lines suggesting that the differences in telomere structure in telomerase positive cells require increased levels of these proteins. Characterization of the role of these up-regulated telomeric genes in ALT is ongoing. (15)

Task 3: To establish whether exogenous expression of a subset of these genes can lead to telomerase independent immortalization of primary HOSE cells. While completing identification of differentially expressed genes, we also initiated a candidate factor approach to characterize the role of tumor suppressor proteins in ALT. This line of investigation has been fruitful with respect to two tumor suppressor proteins important in ovarian cancer, p53 and BRCA1.

Current data suggest that abrogation or mutation of p53 facilitates activation of the ALT pathway (6, 16-18). In addition to its functions responding to DNA damage, p53 also acts to suppress recombination (19-21), independent of transactivation activity (22, 23), raising the possibility that p53 might regulate ALT via its role as a regulator of recombination. To test the role of p53 in ALT we utilized inducible alleles of human p53 (15, 24-26). We found that expression of transactivation incompetent p53 inhibits DNA synthesis in ALT cell lines, but does not affect telomerase positive cell lines (27). Conditional expression of p53 in clonal cell lines resulted in ALT-specific, transactivation-independent growth inhibition, due in part to perturbation of S phase (27). Utilizing chromatin immunoprecipitation (ChIP) assays, we demonstrate that p53 is associated with the telomeric complex in ALT cells (27). Furthermore, inhibition of DNA synthesis in ALT cells by p53 requires an intact ability to bind to telomeric DNA (27). We propose that p53 causes transactivation-independent growth inhibition of ALT cells by suppressing telomeric recombination. Thus, tumors containing p53 mutation may be more likely to activate ALT when challenged with telomerase inhibition. Current efforts in the lab are aimed at testing this hypothesis.

ALT cells are characterized by a specialized version of the promyelocytic leukemia nuclear body, ALT-associated PML bodies (APBs) (8). BRCA1 co-localizes with and is required for DNA replication within APBs (28). We found using chromatin immunoprecipitation that BRCA1 is associated with a complex that contains telomeric DNA in cells that use ALT. The amino terminal 1100 amino acids of BRCA1, spanning the RAD50 interacting domain and the DNA binding domain, are sufficient to permit BRCA1 focus formation and to target BRCA1 to sub-nuclear domains that contain APBs. However, co-localization of BRCA1 with APBs requires the carboxy-terminus of BRCA1. A domain of BRCA1 containing phosphorylation sites for checkpoint kinases enhanced BRCA1 focus formation, but did not affect the frequency of targeting to APBs. Using phospho-specific antibodies we found that BRCA1 associated with APBs is phosphorylated on serine residues 1457 and 1524, targeted by the ATR and ATM checkpoint kinases (29). We are currently determining the role these checkpoint pathways play in formation and maintenance of APBs.

Key Research Accomplishments

- Completed SSH and fabrication of all custom arrays proposed.

- Developed high density oligonucleotide microarrays to establish expression profiles in ALT and telomerase positive cells.
- Characterized 50 clinical ovarian tumors for mechanism of telomere maintenance.
- Created and characterized ALT cell lines in which telomerase has been reconstituted.
- Carried out hybridization and analysis of array data.
- Identified a panel of genes that were differentially expressed in HOSE cells upon immortalization and telomere stabilization.
- Identified a subset of these genes that were differentially expressed in ALT HOSE cells versus telomerase positive HOSE cells.
- Initiated validation of differentially expressed genes by RT-PCR and real time PCR.
- Completed characterization of 50 clinical ovarian tumors for telomerase activity.
- Completed characterization of 50 clinical ovarian tumors for telomere length.
- Developed technique to detect APBs in tumor sections.
- Developed techniques to microdissect tumor tissue from frozen sections and isolated RNA.
- Developed techniques to amplify RNA from microdissected tissues for microarrays and real time PCR.
- Derived human ovarian surface epithelium (HOSE) cells that stably express hTERT.
- Demonstrated that hTERT expression in a ALT cells did not revert their ALT phenotype.
- Reported that BRCA1 co-localizes with and is required for DNA replication within APBs.
- Found that BRCA1 is associated with a complex that contains telomeric DNA in cells that use ALT.
- Demonstrated for the first time that p53 is associated with the telomeric complex in ALT cells.
- Demonstrated that p53 negatively regulates telomere maintenance by the ALT pathway.

Reportable Outcomes(including Bibliography of Publications)

1. Frolov, A., Pan, Z-Z., Broccoli, D., Vanderveer, L., Auersperg, N., Lynch, H., Daly, M., Hamilton, T., Godwin, A.K. Identification of ovarian cancer-associated genes using a HOSE cell transformation model. The Ninth Annual SPORE Workshop, Washington, D.C., p171, July 2001, oral presentation.
2. Broccoli, D., and Godwin, A.K. Telomere length changes in Human Cancer. In Methods in Molecular Biology: The Molecular Analysis of Cancer. Eds J. Boulton & C. Fidler. The Humana Press, pp. 271-278, 2001.
3. Grobely, J.V., Lynch, H., Godwin, A.K., Broccoli, D. *BRCA1* Mutations and Telomerase Activity in Breast and Ovarian Tumors. Telomeres and Telomerase, Cold Spring Harbor, March 28-April 1, 2001.

4. Caslini, C., Frolov, A., Godwin, A.K., Broccoli, D. Gene expression profiling of ALT-positive human ovarian surface epithelial cells reconstituted for telomerase activity. The Seventh Annual Postdoctoral Research Conference, Philadelphia, 2002.
5. Caslini, C., Carlisle, A.J., Godwin, A.K., Broccoli, D. BRCA1 binding to telomeric complex in ALT positive cell lines. Telomeres and Telomerase, Cold Spring Harbor, NY, 2003 (Oral Presentation)
6. Caslini, C., Carlisle, A.J., Godwin, A.K., Broccoli, D. (2003) BRCA1 binding to telomeric complex in ALT positive cell lines. The Seventh Annual Postdoctoral Research Conference, Philadelphia, 2002.
7. Ochs, M. and Godwin, A.K. Microarrays in cancer: research and applications. *BioTechniques*, 34:4-15, 2003.
8. Razak ZR, Varkonyi RJ, Kulp-McEliece M, Caslini C, Testa JR, Murphy ME, Broccoli D. p53 differentially inhibits cell growth depending on the mechanism of telomere maintenance. *Mol Cell Biol*. 24:5967-77, 2004.

Conclusions

The presence of telomerase activity in the majority of tumors and the absence of activity in most human somatic cells has made telomerase an attractive target for cancer therapeutics. Telomerase inhibition can arrest the growth of tumor cells both *in vivo* and *in vitro*. Although these approaches deserve close attention, the presence of telomerase-independent mechanisms for telomere maintenance should not be ignored. Tumors using a telomerase independent mechanism, i.e., ALT (Alternative Lengthening of Telomeres), to maintain telomeric arrays would most likely be refractory to treatment with telomerase inhibitors. Likewise, the ALT pathway represents a salvage pathway that may be activated in tumors in order to overcome therapeutic effects of telomerase inhibitors. It is our hypothesis that identification of the genes that contribute to telomerase independent telomere maintenance in human cells would allow for the development of strategies to combat growth of a significant percentage of ovarian tumors and/or may suggest strategies for prevention. We have made substantial progress towards identifying gene expression changes relevant to ALT and malignant transformation of the ovarian surface epithelial cells. Functional validation of the role of these candidates in malignant transformation of ovarian surface epithelial cells and/or ALT is ongoing. Those genes that are determined to be important in these processes will represent new targets for diagnosis and therapy.

References

1. Testa, J. R., Getts, L. A., Salazar, H., Liu, Z., Handel, L. M., Godwin, A. K., and Hamilton, T. C. Spontaneous transformation of rat ovarian surface epithelial cells results in well to poorly differentiated tumors with a parallel range of cytogenetic complexity. *Cancer Res*, 54: 2778-2784, 1994.
2. Rose, G. S., Tocco, L. M., Granger, G. A., DiSaia, P. J., Hamilton, T. C., Santin, A. D., and Hiserodt, J. C. Development and characterization of a clinically useful animal model of epithelial ovarian cancer in the Fischer 344 rat. *Am J Obstet Gynecol*, 175: 593-599, 1996.
3. Godwin, A. K., Testa, J. R., Handel, L. M., Liu, Z., Vanderveer, L. A., Tracey, P. A., and Hamilton, T. C. Spontaneous transformation of rat ovarian surface epithelial cells: association

- with cytogenetic changes and implications of repeated ovulation in the etiology of ovarian cancer. *J Natl Cancer Inst*, 84: 592-601, 1992.
4. Grobelny, J. V., Godwin, A. K., and Broccoli, D. ALT-associated PML bodies are present in viable cells and are enriched in cells in the G2/M phase of the cell cycle. *Journal of Cell Science*, 113: 4577-4585, 2000.
 5. Shay, J. W. and Bacchetti, S. A survey of telomerase activity in human cancer. *Eur. J. Cancer*, 33: 787-791, 1997.
 6. Bryan, T. M., Englezou, A., Gupta, J., Bacchetti, S., and Reddel, R. R. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.*, 14: 4240-4248, 1995.
 7. Bryan, T. M., Englezou, A., Dalla-Pozza, L., Dunham, M. A., and Reddel, R. R. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.*, 3: 1271-1274, 1997.
 8. Yeager, T., Neumann, A., Englezou, A., Huschtscha, L., Noble, J., and Reddel, R. Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Res*, 59: 4175-4179, 1999.
 9. Ulaner, G. A., Huang, H. Y., Otero, J., Zhao, Z., Ben-Porat, L., Satagopan, J. M., Gorlick, R., Meyers, P., Healey, J. H., Huvos, A. G., Hoffman, A. R., and Ladanyi, M. Absence of a telomere maintenance mechanism as a favorable prognostic factor in patients with osteosarcoma. *Cancer Res*, 63: 1759-1763, 2003.
 10. Montgomery, E., Argani, P., Hicks, J. L., DeMarzo, A. M., and Meeker, A. K. Telomere lengths of translocation-associated and nontranslocation-associated sarcomas differ dramatically. *Am J Pathol*, 164: 1523-1529, 2004.
 11. van Steensel, B., Smogorzewska, A., and de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell*, 92: 401-413, 1998.
 12. Broccoli, D., Smogorzewska, A., Chong, L., and de Lange, T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nature Gen.*, 17: 231-235, 1997.
 13. Karlseder, J., Broccoli, D., Dai, Y., Hardy, S., and de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science*, 283: 1321-1325, 1999.
 14. Cook, B. D., Dynek, J. N., Chang, W., Shostak, G., and Smith, S. Role for the related poly(ADP-Ribose) polymerases tankyrase 1 and 2 at human telomeres. *Mol Cell Biol*, 22: 332-342, 2002.
 15. Jimenez, G. S., Nister, M., Stommel, J. M., Beeche, M., Barcarse, E. A., Zhang, X. Q., O'Gorman, S., and Wahl, G. M. A transactivation-deficient mouse model provides insights into Trp53 regulation and function. *Nat Genet*, 26: 37-43, 2000.
 16. Opitz, O. G., Suliman, Y., Hahn, W. M., Harada, H., Blum, H. E., and Rustig, A. K. Cyclin D1 overexpression and p53 inactivation immortalize primary oral keratinocytes by a telomerase-independent mechanism. *J Clin Invest*, 108: 725-732, 2001.
 17. Sood, A. k., Coffin, J., Sarvenaz, J., Buller, R. E., Hendrix, M. J. C., and Klingelutz, A. p53 null mutations are associated with a telomerase negative phenotype in ovarian carcinoma. *Cancer Biol Ther*, 1: 511-517, 2002.
 18. Henson, J. D., Neumann, A. A., Yeager, T. R., and Reddel, R. R. Alternative lengthening of telomeres in mammalian cells. *Oncogene*, 21: 598-610, 2002.
 19. Sturzbecher, H. W., Donzelmann, B., Henning, W., Knippschild, U., and Buchhop, S. p53 is linked directly to homologous recombination processes via RAD51/RecA protein interaction. *EMBO J*, 15: 1992-2002, 1996.
 20. Saintigny, Y., Rouillard, D., Chaput, B., Soussi, T., and Lopez, B. S. Mutant p53 proteins stimulate spontaneous and radiation-induced intrachromosomal homologous recombination independently of the alteration of the transactivation activity and of the G1 checkpoint. *Oncogene*, 18: 3553-3563, 1999.

21. Mekeel, K. L., Tang, W., Kachnic, L. A., Luo, C. M., DeFrank, J. S., and Powell, S. N. Inactivation of p53 results in high rates of homologous recombination. *Oncogene*, 14: 1847-1857, 1997.
22. Dudenhoffer, C., Kurth, M., Janus, F., Deppert, W., and Wiesmuller, L. Dissociation of the recombination control and the sequence specific transactivation function of p53. *Oncogene*, 18: 5773-5784, 1999.
23. Boehden, G. S., Akyuz, N., Roemer, K., and Wiesmuller, L. p53 mutated in the transactivation domain retains regulatory functions in homology-directed double-strand break repair. *Oncogene*, 22: 4111-4117, 2003.
24. Michalovitz, D., Halevy, D., and Oren, M. Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53. *Cell*, 62: 671-680, 1990.
25. Martinez, J., Georgoff, I., Martinez, J., and Levine, A. J. Cellular localization and cell cycle regulation by a temperature-sensitive p53 protein. *Genes Dev*, 5: 151-159, 1991.
26. Lin, J., Chen, J., Elenbaas, B., and Levine, A. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. *Genes Dev*, 8: 1235-1246, 1994.
27. Razak, Z. R., Varkonyi, R. J., Kulp-McEliece, M., Caslini, C., Testa, J. R., Murphy, M. E., and Broccoli, D. p53 differentially inhibits cell growth depending on the mechanism of telomere maintenance. *Mol Cell Biol*, 24: 5967-5977, 2004.
28. Wu, G., Jiang, X., Lee, W. H., and Chen, P. L. Assembly of functional ALT-associated promyelocytic leukemia bodies requires Nijmegen Breakage Syndrome 1. *Cancer Res*, 63: 2589-2595, 2003.
29. Gatei, M., Zhou, B. B., Hobson, K., Scott, S., Young, D., and Khanna, K. K. Ataxia telangiectasia mutated (ATM) kinase and ATM and Rad3 related kinase mediate phosphorylation of Brca1 at distinct and overlapping sites. In vivo assessment using phospho-specific antibodies. *J Biol Chem*, 276: 17276-17280., 2001.

List of Personnel

Dominique Broccoli – Principal Investigator
 Andrew Godwin – Co-Investigator
 Andrey Frolov – Postdoc Associate
 Robert Varkonyi – Science Tech II
 Natalya Frolova – Science Tech II
 Corrado Caslini – Postdoc Associate

Appendices

1. Broccoli, D., and Godwin, A.K. Telomere length changes in Human Cancer. In *Methods in Molecular Biology: The Molecular Analysis of Cancer*. Eds J. Boulton & C. Fidler. The Humana Press, pp. 271-278, 2001.
2. Razak ZR, Varkonyi RJ, Kulp-McEliece M, Caslini C, Testa JR, Murphy ME, Broccoli D. p53 differentially inhibits cell growth depending on the mechanism of telomere maintenance. *Mol Cell Biol*. 24:5967-77, 2004.
3. Ochs, M. and Godwin, A.K. Microarrays in cancer: research and applications. *BioTechniques*, 34:4-15, 2003.

Telomere Length Changes in Human Cancer

Dominique Broccoli and Andrew K. Godwin

1. Introduction

Telomere length is now known to be directly responsible for limiting the capacity of cellular division in a number of human cell types (1,2). Comparison of telomere length in tumors and matched normal tissue from the same individual has indicated that the telomere repeat array is often shorter in tumors than in adjacent untransformed tissue (3–5) (Fig. 1). Changes in telomere length during human tumorigenesis are believed to reflect loss of terminal sequences resulting from the end replication problem during the cell divisions required for tumor formation. Although telomeres are often shorter in tumors than in matched normal tissue from the same individual, these telomeres are usually stabilized at a new length setting by the activation of the telomere maintenance enzyme, telomerase (6,7).

Telomere length is usually determined by Southern analysis of terminal restriction fragments (TRFs). This is technically the most simple protocol for visualizing telomere arrays. However, difficulties in interpreting the telomere length data generated from this technique arise owing to the heterogeneity in the number of telomeric repeats present at individual chromosome ends. In contrast to the discrete bands usually resulting from Southern analysis, telomeric signals appear as a smear of hybridization. The intensity of the signal is biased toward the larger telomeric fragments because these fragments contain a greater number of target sequences for hybridization. In addition, the TRFs being visualized via this method represent not only the canonical TTAGGG repeats but a variable amount of subtelomeric sequences (Fig. 2).

Recently, two additional methods for ascertaining telomere length have been developed. First, hybridization of TRFs can be done in solution with an oligonucleotide probe complementary to the single-stranded protrusion present at

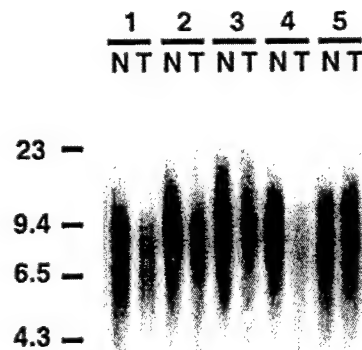


Fig. 1. Analysis of telomere length changes in breast tumors. TRFs are from normal (N) and tumor (T) tissue from five individuals. Genomic DNA was digested with *Hinf*I and visualized following Southern hybridization with the telomeric oligonucleotides (TTAGGG)₄ and (CCCTAA)₄. Note that TRFs are usually shorter in tumors than in adjacent normal tissue from the same individual.

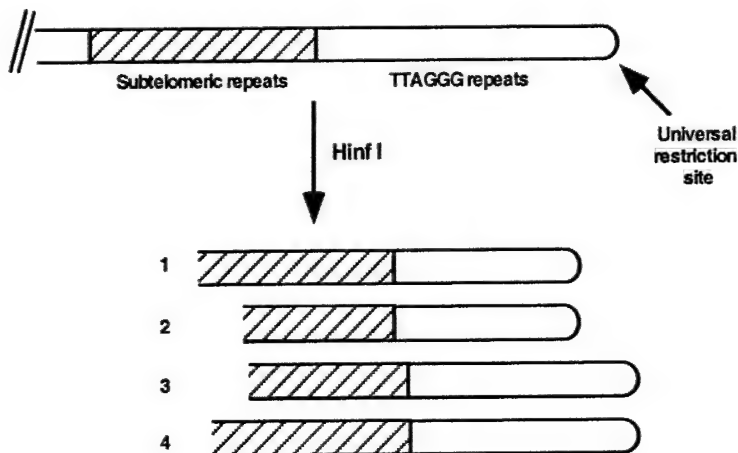


Fig. 2. Telomeric arrays appear as a smear rather than as discrete bands following Southern analysis owing to heterogeneity in the amount of subtelomeric sequences (1 vs 2 and 3 vs 4) and in the number of TTAGGG repeats present at each chromosome end (1 vs 4 and 2 vs 3).

the 3' ends of chromosomes (8,9). This method reduces heterogeneity in signal intensity owing to variable amounts of target sequence, allowing quantitation of telomere length to be somewhat simplified. However, there are several of problems associated with this technique, including the necessity of running hot

gels and the care to determine that one's DNA samples do not become denatured. Second, telomere length can be determined by *in situ* hybridization using peptide nucleic acid-based probes (10). This technique permits accurate determination of telomere length based on signal intensity and allows investigation of telomere length on an individual chromosome basis. However, this technique requires the generation of metaphase spreads from actively growing cultures, and data collection is onerous. Thus, peptide nucleic acid-based *in situ* hybridization is not yet suitable for analysis of telomere length in large numbers of human tumor samples.

This chapter describes the analysis of human telomere length in human tumors using Southern hybridization of TRFs. This is the technique that is most widely used in the field of telomere dynamics, and where possible, we have optimized the use of commercially available reagents and kits to minimize interlaboratory variation.

2. Materials

2.1. Preparation of Genomic DNA

1. Matched samples from tumor and normal tissue.
2. 10X PBSA: 1.37 M NaCl, 26.8 mM KCl, 106 mM Na₂HPO₄, 14.7 mM K₂H₂PO₄.
3. TNE lysis Buffer: 0.5 M Tris-HCl, pH 8.9, 10 mM NaCl, 15 mM EDTA.
4. Proteinase K (10 mg/mL).
5. Phenol (saturated with 0.1 M Tris/0.2% β -mercaptoethanol, pH 8.0).
6. 3 M Sodium acetate, pH 5.7.
7. Absolute and 70% ethanol.
8. T₁₀E₂₀: 10 mM Tris-HCl/20 mM EDTA, pH 7.5.
9. RnaseA (10 mg/mL).
10. 10% Sodium dodecyl sulfate (SDS).
11. Phenol/chloroform/isoamyl alcohol (25:24:1) (Fisher).
12. T₁₀E₁: 10 mM Tris-HCl/1 mM EDTA, pH 8.0.
13. Centrifuge.
14. Polypropylene tubes (15 and 50 mL).
15. Rotator.

2.2. Digestion and Quantitation of DNA Samples

1. *Hinf*I restriction endonuclease and reaction buffer (New England Biolabs).
2. Incubator.
3. Fluorometer (Bio-Rad, Hercules, CA).
4. DNA quantitation kit (Bio-Rad) containing 10X TEN (100 mM Tris-HCl, 2 M NaCl, 10 mM EDTA, pH 7.4), 1 mg/mL of calf thymus DNA, and 10 mg/mL of Hoechst 33258 (bisbenzamide). Alternatively, components for quantitation can be purchased separately.
5. Cuvets.
6. Eppendorf tubes.

2.3. Southern Blotting

1. Agarose.
2. Ethidium bromide (10 mg/mL).
3. 5X Tris-borate EDTA (TBE) buffer.
4. DNA molecular weight standards (New England Biolabs).
5. Agarose gel electrophoresis chamber and power supply.
6. 0.25 *N* HCl.
7. Denaturing solution: 0.5 *M* NaOH, 1.5 *M* NaCl.
8. Neutralizing solution: 1 *M* Tris-HCl, pH 7.5, 1.5 *M* NaCl.
9. Hybond N (Amersham).
10. Stratalinker (Stratagene).
11. Transfer apparatus (capillary or electroblot).

2.4. Probe Preparation and Hybridization

1. Hybridization mix: 0.5 *M* NaH₂PO₄, 10% BSA, 7% SDS.
2. Filter unit (500-mL, 0.45- μ m) (Nalgene).
3. Hybridization setup (water bath, oven, and so on).
4. ProbeQuant G-50 microcolumns (Pharmacia).
5. T4 kinase and 5X Forward Reaction Buffer (Gibco-BRL).
6. γ^{32} P-ATP (3000 Ci/mmol).
7. (TTAGGG)₄ and (CCCTAA)₄ oligonucleotides (100 ng/ μ L).
8. TES: 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, pH 8.0, 0.1% SDS.
9. Pyrex dish or tupperware.
10. 4X Saline sodium citrate (SSC), 0.1% SDS.
11. Autoradiographic film or Phosphorimager cassette.

3. Methods

3.1. Preparation of Genomic DNA (see Note 1)

It is essential that tissue samples are pathologically evaluated to minimize cross contamination of tumor tissue with normal tissue. Tissues should be flash frozen and stored at -70°C until use.

1. Rinse approx 1 g (1 cm³) of frozen tissue with 1X PBSA.
2. Decant the wash and finely mince the specimen (~1 g or 1 cm³) with a razor or scalpel blade.
3. Place the minced tissue into a 50-mL polypropylene tube and add 10 mL of TNE lysis buffer supplemented with 500 μ g/mL of proteinase K and 1% SDS.
4. Parafilm the cap and incubate overnight at 37°C with constant rocking.
5. Add an equal volume of phenol saturated with 0.1 *M* Tris/0.2% β -mercapto-ethanol (pH 8.0), and mix gently at room temperature for 10–15 min.
6. Centrifuge at 1400g for 10 min.
7. Remove the aqueous phase (upper layer) with a wide-bored pipet (e.g., 25-mL disposable pipet), and transfer to a fresh 50-mL polypropylene tube. Repeat the

phenol extraction as described previously until the interphase between the aqueous and organic layers is clear.

8. Transfer the final aqueous phase to a fresh tube. This solution will be very viscous, and care should be taken to slowly and gently remove the aqueous phase to limit the contamination with the phenol solution.
9. Add 0.1 vol of 3 M sodium acetate (pH 5.7) and 2 vol of cold absolute ethanol.
10. Precipitate the DNA by rotating gently for 5–10 min. The genomic DNA will precipitate into a white stringy clump. Loosely spool strands of DNA around a glass pipet or a yellow pipet tip and transfer to a fresh 50-mL tube.
11. Wash the spooled DNA by rotating the DNA in 25–35 mL of 70% EtOH for 5–10 min at room temperature.
12. Air-dry the DNA and place in a 15-mL polypropylene tube (Falcon 2059). Resuspend the DNA in 3 mL of T₁₀E₂₀ by rotating overnight at 4°C.
13. Add 100 µg/mL of RNase A (10 mg/mL stock) and incubate at 37°C for 30 min.
14. Add proteinase K and SDS to final concentrations of 200 µg/mL and 1%, respectively. Incubate for 1 h at 48°C.
15. Extract twice with equal volumes of phenol/chloroform/isoamyl alcohol (25:24:1) as described in steps 5–8.
16. Add 0.1 vol of 3 M sodium acetate, 2 vol of cold absolute EtOH to the aqueous phase, and spool the DNA as before.
17. Wash the DNA with 70% ethanol.
18. Air-dry the DNA and then vacuum dry the pellet.
19. Resuspend the dried DNA in 1 to 2 mL of T₁₀E₁ by rotating overnight at 4°C.
20. Store the DNA samples at 4°C. The concentration of the sample may be determined using a spectrophotometer at this time, but because the viscosity of the solution makes accurate pipetting difficult, these measurements are not very reliable.

3.2. Digestion and Quantitation of DNA Samples

1. Digest 5–10 µL of the DNA sample (no more than 5 µg) in a 50-µL reaction overnight at 37°C with 25U of *Hinf*I. The DNA should be pipetted using sawed-off tips to minimize shearing and facilitate pipetting. Incubation should be carried out in an incubator rather than a water bath to minimize condensation of the reaction on the lid of the Eppendorf tube.
2. Briefly spin the tubes to collect the reaction volumes at the bottom of the tube.
3. Quantitate the amount of DNA in each reaction using the Bio-Rad fluorometer (or similar equipment) and the DNA quantitation kit (Bio-Rad).
4. Prepare sufficient 1X TEN (2 mL/sample to be quantitated).
5. Add Hoechst 33258 to a final concentration of 1 µg/mL.
6. Zero the fluorometer using 2 mL of the TEN/dye solution.
7. Calibrate the fluorometer by adding 5 µL of 100 µg/mL calf thymus DNA (equivalent to a total of 500 ng of DNA).
8. Check the calibration by adding 10 µL of 10 µg/mL calf thymus DNA to 2 mL of TEN/dye solution.

9. Determine the concentration of DNA in each digest by adding 2 μL of the reaction to 2 mL of TEN/dye solution.
10. Calculate the volumes required to load 1 μg to 2.5 μg of DNA/lane on an agarose gel.

3.3. Southern Blotting

1. Pour a $20 \times 20 \text{ cm}^2$ agarose gel. The agarose mix is composed of 0.7% agarose in 0.5X TBE buffer supplemented with 1 $\mu\text{g}/\text{mL}$ of ethidium bromide.
2. Load equal amounts of each DNA sample on the gel. We add 0.1 vol of loading dye to each sample. Also include molecular weight markers (*see Note 2*).
3. Run the gel at 30 V until the dye front has entered the gel. The gel can then be turned up to 80 V or continued running at 30 V. The gel should be run for a total of 700–1000 Vh.
4. Check the gel by observing on an ultraviolet light box. The gel should run until the 2-kb molecular weight marker is at the bottom. The majority of the genomic DNA should have run off the gel since *HinfI* is a frequent cutter in bulk genomic DNA with resulting average sized fragments <2 kb. The intensity of ethidium bromide staining should appear equal for all lanes.
5. Take a picture of the gel with a ruler next to the molecular weight markers.
6. Incubate the gel for 15 min at room temperature with 0.25 N HCl on a rotator. The volume should be sufficient to cover the gel completely.
7. Treat the gel twice for 20 min each in denaturing solution on a rotator.
8. Treat the gel twice for 30 min each in neutralizing solution on a rotator.
9. Rinse the gel with ddH_2O .
10. Transfer the DNA to a Hybond N membrane by capillary transfer or using an electroblot apparatus. After transfer is complete, mark the location of the wells.
11. Crosslink the DNA to the membrane by treating in a Statalinker (Stratagene) at 1200 mJ (the “autocrosslink” setting). The membrane can be stored for an unlimited time or immediately be hybridized to detect the TRFs.

3.4. Probe Preparation and Hybridization

The hybridization mix is made in 500-mL aliquots and will last for months at room temperature. The solution should be incubated at 50°C overnight to ensure complete dissolution of the SDS and BSA. The following morning, the hybridization mix is filtered through a $0.45\text{-}\mu\text{m}$ filter unit. This step eliminates “speckling” appearing on the films. Where possible, the procedures below utilize commercially available reagents to eliminate variation from experiment to experiment.

1. Prehybridize the filter at 55°C in 10 mL hybridization mix for at least 30 min at 55°C .
2. While the filter is prehybridizing, label the oligonucleotides. We use a 1:1 mixture of the two oligonucleotides (a total of 200 ng of oligonucleotide) (*see Note 3*).
3. End label the oligonucleotides using T4 kinase and 20 pmol (12 μL of 3000 Ci/mmol) of $\gamma^{32}\text{P}$ -ATP for 45 min to 1 h at 37°C in a final volume of 20 μL (4 μL of 5X forward reaction buffer, 1 μL of each oligonucleotide [100 ng/ μL], 2 μL of T4 kinase, 12 μL of $\gamma^{32}\text{P}$ -ATP).

4. Stop the reaction by adding 30 μ L of TES to the reaction to achieve a final volume of 50 μ L.
5. Remove the unincorporated radionuclotide from the reaction using ProbeQuant G-50 microcolumns (Pharmacia).
6. Determine the efficiency of labeling in a crude manner using a Geiger counter. The probe should be labeled such that the counter is saturated on the most sensitive setting when the monitor is held approx 1 in. from the tube.
7. Add the probe to 7 mL of hybridization mix. Replace the prehybridization solution with the hybridization solution and incubate at 55°C overnight.
8. Wash the filter twice for 20 min each in 4X SSC/0.1% SDS at room temperature. Use a Geiger counter to monitor the filter. There should be very few counts remaining in the center of the filter and background levels of radiation at the top and bottom of the filter. If necessary, wash the filter an additional time in 4X SSC/0.1% SDS at 55°C.
9. Wrap the filter in plastic wrap and expose to autoradiographic film or a Phosphorimager cassette. The exposure time and use of intensifying screens varies with experiments and efficiency of hybridization and can range from 2 h with two screens at -70°C to 3 d with two screens at -70°C.

4. Notes

1. Genomic DNA can be prepared using a number of commercially available kits. We have had mixed results using kits and recommend determining the quality of DNA prepared in this manner by running 1 μ L of the sample on a gel prior to digestion. Up to 50% of the genomic DNA isolated using kits may be degraded and therefore unsuitable for this type of analysis.
2. Internal telomeric fragments of approx 2 and 2.3 kb act as internal controls for loading and the integrity of DNA.
3. Because of the highly repetitive nature of telomeres, hybridization to these sequences is occasionally uneven. This is owing to insufficient quantities of probe sequence relative to target sequence, i.e., unsaturated conditions. The amount of oligonucleotide suggested in the protocol is usually sufficient to prevent this problem. However, in the event of uneven hybridization patterns, 100–200 ng of unlabeled oligonucleotides should be added to the hybridization mix.

References

1. Vaziri, H. and Benchimol, S. (1998) Reconstitution of telomerase activity in normal human cells leads to elongation of telomeres and extended replicative life span. *Curr. Biol.* **8**, 279–282.
2. Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., et al. (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science* **279**, 349–352.
3. de Lange, T. (1998) Telomeres and senescence: ending the debate. *Science* **279**, 334,335.

4. de Lange, T., Shiue, L., Myers, R. M., Cox, D. R., Naylor, S. L., Killery, A. M., and Varmus, H. E. (1990) Structure and variability of human chromosome ends. *Mol. Cell. Biol.* **10**, 518–527.
5. Hastie, N. D., Dempster, M., Dunlop, M. G., Thompson, A. M., Green, D. K., and Allshire, R. C. (1990) Telomere reduction in human colorectal carcinoma and with ageing. *Nature* **346**, 866–868.
6. Shay, J. W. and Wright, W. E. (1996) The reactivation of telomerase activity in cancer progression. *Trends Genet.* **12**, 129–131.
7. Shay, J. W. and Bacchetti, S. (1997) A survey of telomerase activity in human cancer. *Eur. J. Cancer* **33**, 787–791.
8. Makarov, V., Hirose, Y., and Langmore, J. P. (1997) Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. *Cell* **88**, 657–666.
9. McElligott, R. and Wellinger, R. J. (1997) The terminal DNA structure of mammalian chromosomes. *EMBO J.* **16**, 3705–3714.
10. Lansdorp, P. M., Verwoerd, N. P., van de Rijke, F. M., Dragowska, V., Little, M.-T., Dirks, R. W., et al. (1996) Heterogeneity in telomere length of human chromosomes. *Hum. Mol. Genet.* **5**, 685–691.

Appendix 2

p53 Differentially Inhibits Cell Growth Depending on the Mechanism of Telomere Maintenance

Zaineb R. Abdul Razak,¹ Robert J. Varkonyi,¹ Michelle Kulp-McEliece,^{1†} Corrado Caslini,¹
Joseph R. Testa,² Maureen E. Murphy,³ and Dominique Broccoli^{1*}

Department of Medical Oncology,¹ Human Genetics Program,² and Department of Pharmacology,³ Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111

Received 2 October 2003/Returned for modification 27 November 2003/Accepted 26 March 2004

Telomere stabilization is critical for tumorigenesis. A number of tumors and cell lines use a recombination-based mechanism, alternative lengthening of telomeres (ALT), to maintain telomere repeat arrays. Current data suggest that the mutation of p53 facilitates the activation of this pathway. In addition to its functions in response to DNA damage, p53 also acts to suppress recombination, independent of transactivation activity, raising the possibility that p53 might regulate the ALT mechanism via its role as a regulator of recombination. To test the role of p53 in ALT we utilized inducible alleles of human p53. We show that expression of transactivation-incompetent p53 inhibits DNA synthesis in ALT cell lines but does not affect telomerase-positive cell lines. The expression of temperature-sensitive p53 in clonal cell lines results in ALT-specific, transactivation-independent growth inhibition, due in part to the perturbation of S phase. Utilizing chromatin immunoprecipitation assays, we demonstrate that p53 is associated with the telomeric complex in ALT cells. Furthermore, the inhibition of DNA synthesis in ALT cells by p53 requires intact specific DNA binding and suppression of recombination functions. We propose that p53 causes transactivation-independent growth inhibition of ALT cells by perturbing telomeric recombination.

Telomeres are specialized structures that confer stability to naturally occurring ends of DNA molecules. Telomere stabilization is critical for the unlimited cellular proliferation that is necessary for tumor formation. While most tumors achieve telomere stabilization through the activation of telomerase (48), a subset of tumors utilize a telomerase-independent mechanism termed alternative lengthening of telomeres (ALT) to maintain chromosome termini (7, 8). Telomere length in ALT-positive cell lines is highly heterogeneous, with repeats ranging in size from <5 kb to >20 kb (8). A subset of cells in ALT-positive cell lines also contain large multiprotein complexes in which the telomere binding proteins TRF1 and TRF2 and telomeric DNA colocalize with the promyelocytic leukemia (PML) nuclear body, termed ALT-associated PML bodies (APBs) (65). The PML nuclear body is a multiprotein nuclear structure that has been implicated in the control of a number of cellular processes including apoptosis (41, 47), leading to the suggestion that cells containing APBs might be targeted to undergo apoptosis. However, APB-positive cells incorporate bromodeoxyuridine (BrdU) and, thus, are able to carry out DNA replication (22). In addition, the frequency of cells containing APBs is increased when cultures are enriched for cells in the late S phase or G₂/M phase of the cell cycle (22, 62), suggesting that the formation of APBs is coordinately regulated with the cell cycle.

Studies carried out with *Saccharomyces cerevisiae* indicate that telomerase-independent telomere maintenance occurs via

a recombination-based mechanism (12, 32, 55). Telomere elongation is RAD52 dependent and may occur through a recombination of either telomeric or subtelomeric repeats. It is likely that a recombination-based mechanism also underlies ALT in mammalian cells. Consistent with this hypothesis, immunohistochemical analysis demonstrated that RAD51, RAD52, and the RAD50/MRE11/NBS1 complex colocalize with APBs (65). Studies investigating the fate of a single marked telomere in an ALT-positive cell line also support a recombination-based mechanism (42). In these experiments, rapid changes in the length of the telomere repeat array were observed rather than the gradual changes in size more commonly associated with telomerase activity or telomere loss associated with cell divisions. Furthermore, it has been demonstrated that a unique tag embedded in a single telomere will spread to other telomeres in ALT cell lines, leading to the proposal that telomere elongation occurs through intertelomeric gene conversion (17). However, this only occurs when the tag is flanked by telomeric sequences, suggesting that the recombination event occurs within the TTAGGG repeat array. In contrast, the characterization of the telomeric structure in mouse embryonic stem cells deficient in telomerase suggests that recombination may occur within subtelomeric repeats (38).

The mutation of the tumor suppressor protein p53 has been implicated as a contributing factor for ALT activation. Over 80% of the cell lines that use ALT for telomere maintenance are impaired in the p53 pathway, either due to the expression of viral oncoproteins or through a p53 mutation (26). In one study, all 10 of the cell lines derived from breast fibroblasts of an individual with Li Fraumeni syndrome, carrying a germ line mutation in p53, used ALT for telomere maintenance (8). Similarly, the expression of dominant-negative p53 in conjunc-

* Corresponding author. Mailing address: Department of Oncology, Fox Chase Cancer Center, 7701 Burholme Ave., Philadelphia, PA 19111. Phone: (215) 728-7133. Fax: (215) 728-4333. E-mail: K_Broccoli@fccc.edu.

† Present address: Gwynedd-Mercy College, Gwynedd Valley, PA 19437.

tion with the overexpression of cyclin D1 yielded an ALT-positive immortal cell line (39). Finally, it has recently been reported that ovarian tumors harboring p53 null mutations are more likely to be telomerase negative (50), although it was not established in this study how many of these tumors exhibited the ultralong telomeres and APBs that are characteristic of ALT. While these data suggest that p53 mutation provides a permissive environment for the activation of the ALT pathway of telomere maintenance, it must be pointed out that the majority of p53-compromised cell lines and tumors use telomerase to achieve telomere stabilization. Thus, mutation of the p53 pathway is not sufficient for ALT activation but may be one of several required changes.

Interestingly, from the standpoint of recombination at telomeres, p53 has been found to suppress recombination *in vivo*. For example, cell lines expressing mutant p53 have increased rates of homologous recombination (3, 34, 58). Likewise, the expression of p53 can inhibit the generation of a recombinant chromosome derived from related simian virus 40 (SV40) chromosomes in cell culture (16, 57). p53 also interacts physically with RAD51 and RecA, inhibiting their function (53), and has been found to bind to Holliday junctions and facilitate cleavage of these structures by the Holliday junction resolvases T4 endonuclease VII and T7 endonuclease I *in vitro* (29). The suppression of recombination by p53 does not require its transactivation function (4, 15, 59), suggesting that p53 exerts its effect on recombination independent of changes in gene expression. Because present evidence indicates that ALT involves telomeric recombination, that loss of p53 function is correlated with activation of ALT, and that p53 is involved in suppressing recombination, we sought to determine the role of p53 in the regulation of ALT.

MATERIALS AND METHODS

Cell lines, constructs, and transfection conditions. HIO107, HIO80, HIO120, and HIO114 human ovarian surface epithelium cell lines (1), derived from four different individuals, were maintained in a 1:1 mixture of Media 199 and MCDB-105 medium supplemented with 4% fetal bovine serum and 0.2 IU of pork insulin (Lilly) per ml. The mesothelioma 6 (Meso6) cell line (19) was maintained in RPMI medium with 10% fetal bovine serum. The Wi38VA13/RA2 embryonic lung fibroblast cell line was maintained in Dulbecco's modified Eagle's medium with 10% fetal bovine serum. Temperature-sensitive alleles of p53 (ts-p53) have been described in detail previously. Temperature sensitivity is conferred by a well-characterized mutation of alanine to valine at position 138 (33, 35). The loss of transactivation function was achieved by mutating amino acid 22 from leucine to serine and amino acid 23 from tryptophan to glutamine, resulting in p53(22/23) (31). Both alleles of p53 are inactive when cells are cultured at 39°C and are active when cells are cultured at 32°C. The R273H, R248Q, and R175H mutations were introduced into the p53(22/23) allele by using the QuikChange XL site-directed mutagenesis kit (Stratagene) and the primer pairs R273H-F, 5'-CA GCTTTGAGGTGCATGTTGTGCCTGTCC, and R273H-R, 5'-GGACAGGC ACAACATGCACCTCAAAGCTG; R248Q-F, 5'-GGCGGCATGAACAG AGGCCATCCTC, and R248Q-R, 5'-GAGGATGGGCGCTCTGGTTCATGCC GCC; and R175H-F, 5'-GGAGGTTGTGAGGCACTGCCCCACCATG, and R175H-R, 5'-CATGGTGGGGCAGTGCTCACACCTCC, respectively, as recommended by the manufacturer. The incorporation of the point mutation was confirmed by sequencing. To generate stable lines that express temperature-sensitive wild-type or transactivation-incompetent p53 in HIO107 and HIO114, cells were transfected by electroporation by using a Gene Pulsar II apparatus (Bio-Rad) at 280 V and 975 μ F. For transient transfections, all of the above cell lines were treated with FUGENE 6 reagent (Roche) as recommended by the manufacturer. The cells were subcultured directly onto coverslips 24 h posttransfection and allowed to recover overnight at 39°C. For BrdU analysis, cultures were grown at 39 or 32°C for the 48 h prior to pulsing with 10 μ M BrdU for either 40 min or 24 h.

Indirect immunofluorescence. Cells were grown directly on coverslips and processed for immunofluorescence as described previously (22). The detection of APBs was carried out by using an affinity-purified rabbit polyclonal antibody (FC-08 [22]) raised against a peptide contained in the amino terminal domain of hTRF2 diluted 1:200 and a goat polyclonal antibody against the N terminus of PML (N-19; Santa Cruz) diluted 1:5,000.

Detection of BrdU incorporation was carried out by using a BrdU labeling and detection kit I (Roche) according to the recommendations of the manufacturer. The cells were harvested immediately after BrdU exposure. p53 was detected by using a rabbit polyclonal antibody (FL-393; Santa Cruz) diluted 1:20,000. BrdU was detected by using a mouse monoclonal antibody (diluted 1:50) provided by the manufacturer (Roche). For each cell line, at least 50 cells were scored.

Primary antibodies were detected by using tetramethyl rhodamine isothiocyanate-conjugated donkey anti-rabbit immunoglobulin G (IgG), fluorescein isothiocyanate (FITC)-conjugated donkey anti-goat IgG and FITC-conjugated donkey anti-mouse IgG (Jackson ImmunoResearch). The secondary antibodies did not cross-react. Following incubation with secondary antibodies, the DNA was stained with 0.2 μ g of DAPI (4',6'-diamidino-2-phenylindole) per ml.

Western blot analysis. Protein lysates were prepared by solubilizing 5×10^5 cells in Laemmli sample buffer (Bio-Rad) per μ l. Protein from 2×10^4 cells were separated on sodium dodecyl sulfate (SDS)-10% polyacrylamide gel electrophoresis (PAGE) gels per lane. Proteins were detected by using the following primary antibodies: rabbit polyclonal antibody against p53 (FL-393; Santa Cruz), mouse monoclonal antibody against p21^{WAF1} (F-5; Santa Cruz), and mouse monoclonal antibody against SV40 large T antigen (T-Ag) (Pab108; Santa Cruz). Secondary antibodies used were horseradish peroxidase-linked whole antibody raised in donkey against rabbit IgG and sheep against mouse IgG (Amersham Pharmacia) and were detected by either an ECL kit (Amersham Pharmacia) or SuperSignal West Dura (Pierce).

Viability and apoptosis assay. Viable and apoptotic cells were quantitated by using a Guava personal cytometer and the Guava ViaCount Reagent or the Guava Nexin Kit, respectively, following the manufacturer's specifications. Cells were incubated at 39 or 32°C and harvested every 24 h for up to 96 h. For viability, cells were stained with the ViaCount Reagent. Apoptotic cells were detected by staining with annexin V and 7-amino-actinomycin D (7 AAD) by using the Guava Nexin kit. Both early apoptotic (annexin V-positive) and late apoptotic (annexin V- and 7 AAD-positive) cells were included in the analysis.

FACS analysis. For hydroxyurea (HU) arrests, cells were exposed to 1 mM HU for 22 h prior to harvesting for fluorescence-activated cell sorter (FACS) analysis. For BrdU analysis, cells were pulsed with 10 μ M BrdU for 40 min just prior to harvest. Cells were collected by trypsinization, washed twice with 1 \times phosphate buffered saline (PBS)-2 mM EDTA and fixed in cold 70% ethanol. Following fixation, the DNA was denatured by exposure to 2 M HCl-0.5% Triton X-100 for 30 min at room temperature. The acid was then neutralized with 0.1 M sodium borate, and BrdU was detected by incubating for 30 min at room temperature with FITC-conjugated mouse anti-BrdU (Becton Dickinson). DNA was detected by staining with propidium iodide (50 mg/ml). For dual analysis of BrdU and CD19, the cells were transiently cotransfected with CD19 (37) in conjunction with either a vector or ts-p53TA by using FUGENE 6 reagent, subcultured 24 h posttransfection, and allowed to recover for 4 to 5 h at 39°C. The cells were then shifted to 32°C for an additional 48 h prior to pulsing with 10 μ M BrdU for 40 min. Cells were harvested by trypsinization, washed twice with 1 \times PBS-0.1% bovine serum albumin-0.02% sodium azide, incubated with phycoerythrin-conjugated anti-CD19 (Becton Dickinson) for 1 h at 4°C, and fixed in cold 70% ethanol. After at least 12 h of fixation at 4°C in the dark, the cells were incubated with 1% paraformaldehyde-0.01% Tween 20 in PBS for 30 min, followed by treatment with DNase I for 10 min at room temperature. BrdU was detected by incubating the cells for 30 min at room temperature with FITC-conjugated mouse anti-BrdU (Becton Dickinson). For all analyses, the cells were analyzed by using a Becton-Dickinson FACScan and FlowJo Software.

Chromatin immunoprecipitation assay. Chromatin immunoprecipitation (ChIP) was carried out using the ChIP assay reagents (Upstate Biotechnology Laboratories) essentially as described by the manufacturer. Briefly, 1×10^7 cells were cross-linked for 10 min at room temperature by the addition of formaldehyde to a final concentration of 1%. The reaction was stopped by adding glycine to a final concentration of 0.125 M. Nuclei were isolated, and the chromatin was sonicated 10 times for 10 s each time by using a 2-mm tip and 60% power to generate fragments averaging 1 to 2 kb. GammaBind Plus Sepharose beads (Amersham) were incubated with the chromatin for 1 h at 4°C to preclear the chromatin. The precleared chromatin was incubated overnight at 4°C with the following primary antibodies: goat polyclonal human TRF2 (Imgenex), goat polyclonal p53 (C-19; Santa Cruz), or goat serum IgG (Santa Cruz). Immune complexes were collected on GammaBind Plus Sepharose beads for 2 h at 4°C.

Following washes, the immune complex was eluted from the beads as described by the manufacturer. Cross-linking was reversed by incubation in 0.2 M NaCl for 6 h at 60°C. The resulting DNA was purified by treatment with proteinase K, phenol-chloroform extraction, and ethanol precipitation. The DNA was resuspended in water, and one-half of the sample was interrogated by slot blot hybridization by using oligonucleotides complementary to the telomeric repeats as described previously (5). The filter was stripped and reprobed to detect centromeric alpha satellite (36) as described previously (2). One-quarter of the sample was interrogated by PCR to detect the p53 binding site at position -1.4 kb within the p21^{WAF1} promoter by using the primers 5'ACAGCAGAGGAGAAAGAAGCC3' and 5'ATTACTGACATCTCAGGCTGC3'. PCR was carried out under standard conditions (a 0.2 mM concentration of each deoxynucleoside triphosphate, 1.5 mM MgCl₂, 1× *Taq* buffer, and 0.5 U of *Taq* polymerase) for 40 cycles at an annealing temperature of 61°C to generate a 113-bp product spanning positions -1348 to -1461 of the p21^{WAF1} promoter (GenBank accession number U24170).

RESULTS

p53 inhibits DNA replication in cell lines that use ALT for telomere maintenance. To determine if p53 inhibits the growth of ALT cells, we made use of four ALT cell lines. Three of these were derived from human ovarian surface epithelium from independent individuals (HIO80, HIO107, and HIO120) (1, 22), and the fourth cell line, WI38-VA13/2RA was derived from fetal human lung fibroblasts. In addition, three telomerase-positive cell lines were used. The first is from a human ovarian surface epithelium cell line derived from a fourth individual (HIO114), the second is derived from a mesothelioma (Meso6[19]), and the third is a HeLa cell line. All the HIO cell lines and WI38-VA13 were immortalized with SV40 and contain large T-Ag. Large T-Ag binds to and sequesters p53, preventing it from responding to DNA damage and rendering cells containing this viral oncoprotein null for the p53 pathway. The Meso6 cell line is derived from a malignant mesothelioma and also contains SV40 (19). Finally, in HeLa cells p53 is inactivated by human papillomavirus E6/E7.

The cell lines were transiently transfected with a temperature-sensitive allele of human p53 (ts-p53) containing a mutation in the transactivation domain that impairs the ability of p53 to confer growth arrest or apoptosis in response to DNA damage (27, 31, 44). At 39°C, this transactivation-incompetent allele (ts-p53TA) is in mutant conformation and unable to bind specifically to DNA (33, 35). At 32°C, ts-p53TA is in a wild-type conformation and is able to bind to p53 target sequences. It should be noted that at both 39 and 32°C ts-p53 alleles retain nonspecific DNA binding activity (30, 63), and so phenotypes observed only at 32°C require the specific DNA binding activity of p53.

To quantitate the level of proliferation, the nucleotide analog BrdU was added to transfected cells for the final 40 min of culture in order to label cells in S phase. Indirect immunofluorescence was carried out to detect p53 and BrdU (Fig. 1A). Due to the presence of large T-Ag, all the cells in each culture contain stabilized p53. However, under the staining conditions used here, only the transfected cells overexpressing p53 were positive and were scored (Fig. 1A). Because p53 is overexpressed, the protein is often present in both the cytoplasm and the nucleus. The frequency of cells with high levels of p53 (i.e., that were expressing the exogenous p53 allele) that were also BrdU positive was determined for each cell line at 39 and at 32°C (Fig. 1B, ts-p53TA+). The effect of the different culture temperatures was controlled for by scoring the difference in

the frequency of BrdU-positive cells transfected with empty vector at the two temperatures (Fig. 1B, vector). A ratio of BrdU-positive cells at the two temperatures was then generated, where a value of 1 indicates no difference in the amount of BrdU incorporation when cells are cultured at the two temperatures. The ratio of BrdU incorporation at the two temperatures in cells transfected with the empty vector ranged from 0.4 in the Meso6 cell line (telomerase positive) to 0.9 in the HIO120 cell line (ALT positive), indicating that cell growth was inhibited at 32°C to various extents in the different cell backgrounds (Fig. 1B). The expression of ts-p53TA at the active temperature of 32°C reduced the frequency of BrdU-positive cells in all ALT cell lines tested relative to the ratio generated from cells transfected with empty vector (Fig. 1B), indicating that the expression of ts-p53TA inhibits DNA replication beyond the effects induced by culturing at a lower temperature. In contrast, the expression of ts-p53TA did not have an inhibitory effect on BrdU incorporation in the telomerase-positive cell lines. There was either a similar ratio of BrdU incorporation in cells transfected with ts-p53TA compared to the vector or a higher frequency of BrdU-positive cells when ts-p53TA was expressed.

To assess the inhibition of DNA replication in an independent assay, cells were cotransfected with ts-p53TA and the cell surface marker CD19 (37). Following culture as above, cells were harvested for FACS analysis and stained for BrdU incorporation and CD19 expression. Transfected cells were identified based upon being CD19 positive, and the frequency of BrdU-positive cells in this population at either 39 or 32°C was determined (Fig. 1C and 1D). Again, the cell lines were variably sensitive to growth at 32°C as demonstrated by transfection with empty vector, and the expression of ts-p53TA inhibited DNA replication specifically in the ALT cell lines. Together, these data indicate that p53 inhibits cellular proliferation, as measured by decreased DNA replication, in a transactivation-independent manner specifically in ALT cell lines.

Transactivation-incompetent p53 inhibits growth of cell lines that use ALT for telomere maintenance. To further define the growth inhibitory role of p53 in ALT cells, we generated clonal cell lines that express similar levels of temperature-sensitive human p53 protein (ts-p53WT) or ts-p53TA protein in the HIO107 ALT and HIO114 telomerase-positive parental backgrounds. Both the HIO107 and HIO114 cell lines retain wild-type p53 (A.K. Godwin, unpublished data) which is inactivated by the presence of large T-Ag. The expression levels of p53 are increased in the p53 clonal cell lines relative to the parental cell lines (Fig. 2A and data not shown). This increase in the expression of p53 is sufficient to activate cellular responses, because cells expressing ts-p53WT accumulate p21^{WAF1} protein, a p53-responsive gene, when cultured at 32°C (Fig. 2A). In contrast, p21^{WAF1} does not accumulate in the cell lines expressing ts-p53TA, despite a similar increase in the level of p53 protein relative to the parental cell lines. None of the cell lines accumulated significant levels of p21^{WAF1} at 39°C, when both p53 proteins are in mutant conformation. The lack of p21^{WAF1} accumulation in the cell lines expressing the ts-p53TA allele also indicates that significant levels of the endogenous wild-type p53 are not released from large T-Ag in the presence of the exogenously expressed p53. As expected, expression of ts-p53WT at 32°C resulted in increased levels of

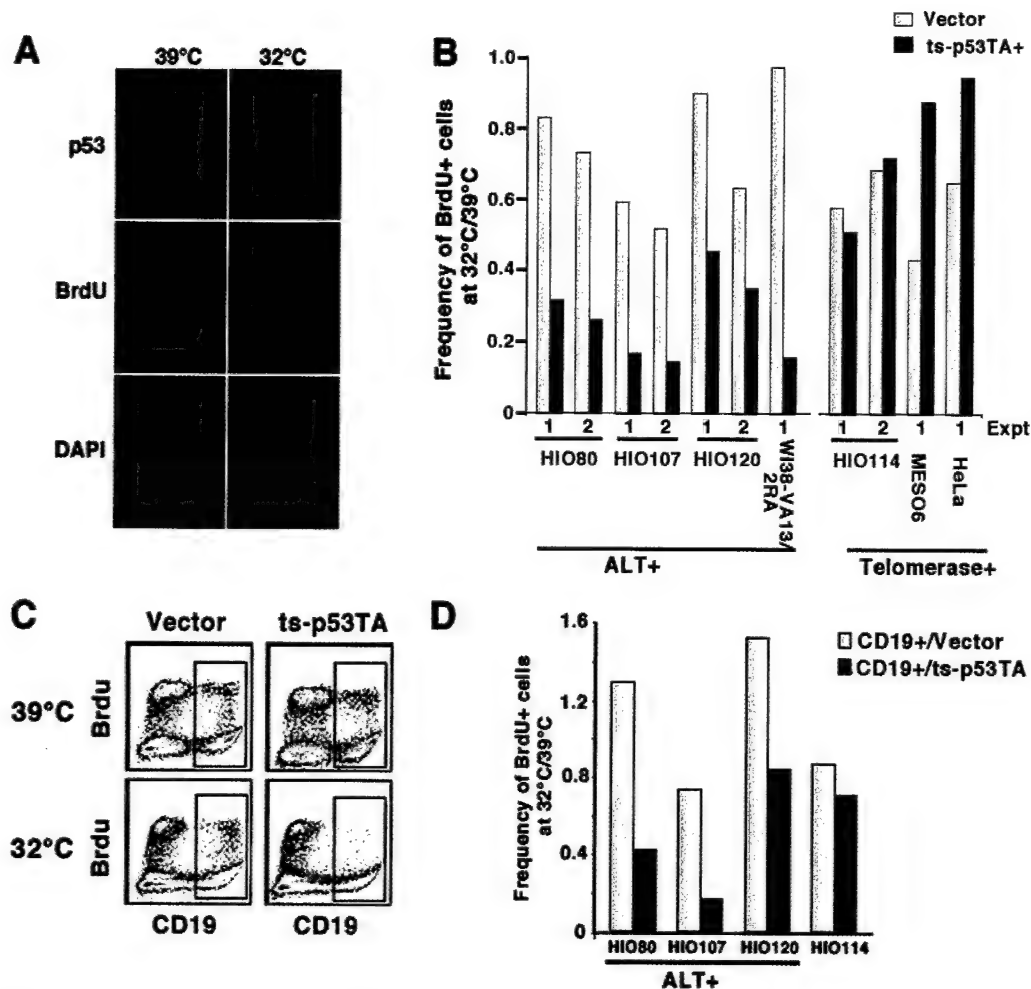


FIG. 1. Expression of ts-p53TA inhibits DNA replication in ALT cell lines. (A) HIO107 cells transiently transfected with ts-p53TA were cultured at either 39 or 32°C for 48 h, pulsed for 40 min with BrdU, and then harvested to detect p53 (red) and BrdU (green). DNA was detected with DAPI. Note that the ts-p53TA-positive cells cultured at 32°C have not incorporated BrdU. Due to overexpression of p53, the protein may be present in both the nucleus and the cytoplasm. (B) Quantitation of data such as those shown in panel A generated from the indicated cell lines. The ALT cell lines expressing ts-p53TA have a reduced frequency of BrdU-positive cells when cultured at 32°C compared to the frequency of BrdU-positive cells transfected with empty vector. (C) FACS analysis of HIO107 cells cotransfected with the CD19 cell surface marker and ts-p53TA. Cells were cultured and pulsed with BrdU as in panel A and then stained to detect BrdU (y axis) and CD19 (x axis). (D) Quantitation of data such as those shown in panel C for the indicated cell lines. The percentage of BrdU-positive cells in the transfected population is determined by gating on the CD19-positive population. The expression of ts-p53TA does not inhibit BrdU incorporation at 32°C in the telomerase-positive cell lines relative to cells transfected with empty vector in either assay.

apoptosis over 72 h (Fig. 2B). The difference in the relative levels of apoptosis following the expression of ts-p53WT in the HIO107 and HIO114 cell lines is due to an approximate two-fold higher level of background apoptosis in the HIO114 cells. The increased apoptosis that occurs following the expression of ts-p53WT at 32°C can be attributed to active p53 as it is not observed when either cell line is grown at 39°C and is not observed at either temperature in the parental cell lines. Consistent with previous reports (43, 44), apoptosis is impaired in both the HIO107-TA and HIO114-TA cell lines relative to the cell lines expressing ts-p53WT (Fig. 2B).

The effects of p53 expression on cell growth were assessed. The HIO107 ALT and HIO114 telomerase-positive parental cell lines proliferate at both temperatures, although growth is slower at 32 than at 39°C. As expected, the expression of

ts-p53WT in either the HIO107 or the HIO114 cell line causes cell death (Fig. 2C). The HIO107-TA ALT cell line also exhibited growth inhibition at 32°C (Fig. 2C), while the HIO114-TA telomerase-positive cell line continued to proliferate. Similar results were obtained with other independent clones of HIO107 and HIO114 expressing ts-p53WT and ts-p53TA (data not shown). These data suggest that p53 specifically inhibits the growth of ALT cells through a transactivation-independent mechanism.

The inhibition of DNA replication occurs rapidly after the expression of ts-p53TA at 32°C. Thus, this phenotype is unlikely to rely upon the gradual telomere attrition that is associated with cellular division in the absence of a mechanism to replicate chromosome termini. However, to establish if expression of the p53 alleles induced rapid alterations in telomeric

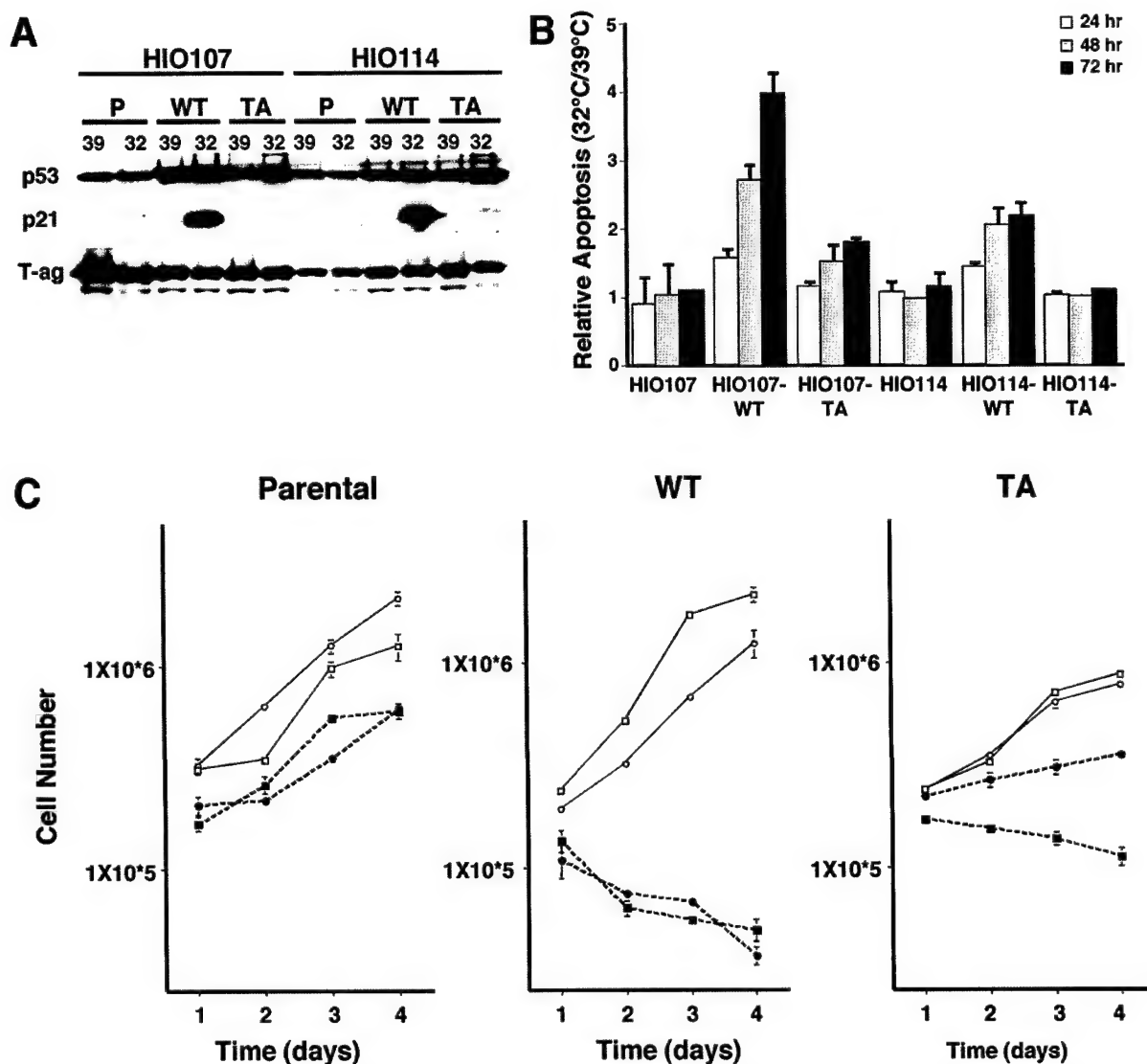


FIG. 2. Characterization of ALT-positive HIO107 or telomerase-positive HIO114 cell lines expressing either wild-type (WT) or transactivation-incompetent (TA) p53. (A) Western blot analysis of whole-cell extracts from the indicated cell lines prepared after 48 h of growth at 39 or 32°C probed to detect p53, p21^{WAF1}, and T-Ag as a loading control. p53-TA does not result in accumulation of p21^{WAF1} protein. P, parental cell line. (B) Frequency of apoptotic cells, detected by 7 AAD and annexin V staining, in the indicated cell lines. Results are shown as the increase (*n*-fold) in apoptosis when cells are cultured at 32°C compared to the level of background apoptosis in the culture at 39°C. Relative to ts-p53WT, the ts-p53TA allele is impaired in its ability to induce apoptosis at the permissive temperature of 32°C. (C) Growth curves of the HIO107 ALT (squares) or HIO114 telomerase-positive (circles) derived cell lines at 39°C (open symbols; solid line) or 32°C (filled symbols; dashed line). Expression of p53-TA causes growth inhibition in the HIO107 ALT background. The results are the summary of experiments done in triplicate. Where they are not visible, the standard error bars are below the resolution of the graph.

structure, we carried out an analysis of terminal restriction fragments. No change in telomere length or in the G-strand overhang was detectable in the HIO107-TA cell line when cultured at 32°C compared to results at 39°C (data not shown). These data indicate that the growth inhibition caused by the ts-p53TA allele is also not due to rapidly occurring global alterations in telomere structure. However, because cellular responses may be triggered by a few dysfunctional telomeres (14, 25), it is unlikely that changes involving a minority of telomeres would be detected.

p53 increases the frequency of ALT-associated PML nuclear bodies. To determine if the expression of p53 had an effect on the ALT pathway, we analyzed the frequency of APBs in which telomeric DNA and proteins colocalize with the PML nuclear body (65). APB-positive cells increase in frequency in both the HIO107-WT and HIO107-TA cell lines when cells are cultured at 32°C, relative to the frequency of APB-positive cells at 39°C (Fig. 3A). In contrast, in the parental HIO107 cell line, cells containing APBs are slightly more frequent at 39°C, suggesting that the increased frequency of APB-containing cells at 32°C is

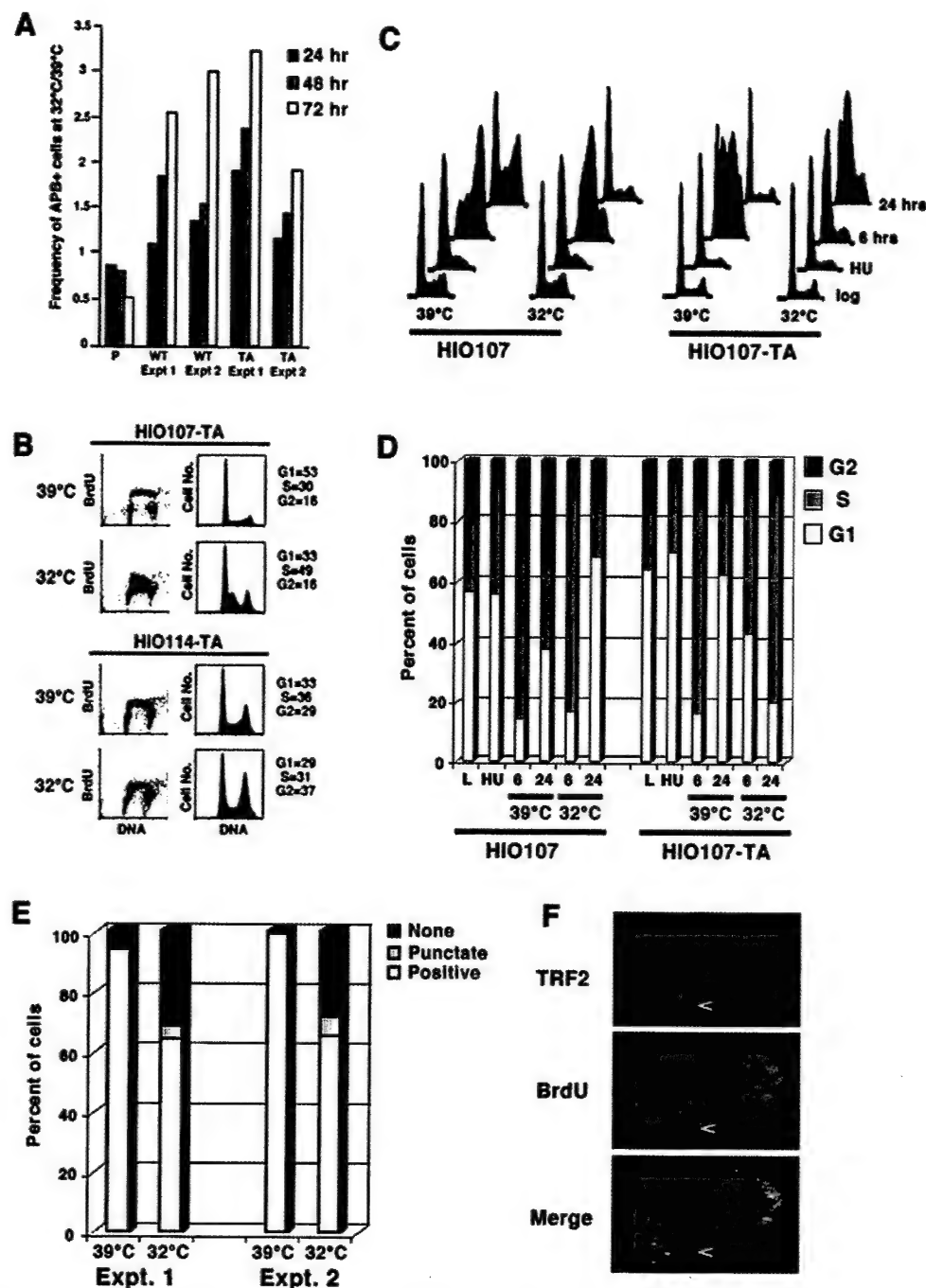


FIG. 3. Expression of ts-p53TA perturbs APB frequency and causes an S phase delay in HIO107 ALT cells. (A) The frequency of APB-positive cells is increased two- to threefold when cell lines expressing either ts-p53WT or ts-p53TA are cultured at 32°C. APBs were detected by costaining the cells with a goat polyclonal antibody against PML and a rabbit polyclonal antibody (FC-08) against TRF2. (B) The HIO107-TA ALT cell line incorporates BrdU, but the cells accumulate in early S phase after 72 h at 32°C, while the HIO114 cell line is unaffected by growth at 32°C. The left panel shows a FACS analysis of BrdU intensity (y axis) versus DNA content (x axis). The right panel shows the cell cycle profiles generated from the data on the left of the figure (y axis, number of events; x axis, DNA content). (C) The parental HIO107 and the HIO107-TA cell lines were arrested by exposure to HU and, following release, cultured at either 32 or 39°C for the indicated time prior to harvest. The presence of active ts-p53TA prevents cells from progressing through S phase following the removal of HU. y axis, number of events; x axis, DNA content. (D) Quantitation of the cell cycle distribution of the populations shown in panel C. HIO107-TA cells accumulate in S phase for up to 24 h, suggesting that S phase is delayed when ts-p53TA is expressed. (E) Quantitation of the pattern of BrdU incorporation in HIO107-TA cells after 72 h of growth at 32°C. The majority of the cells are BrdU positive at 39°C, while at 32°C many of the cells are either BrdU negative or have punctate BrdU staining. (F) At 32°C, some HIO107-TA cells exhibit punctate BrdU staining which colocalizes with the telomeric protein TRF2 (arrow). The other two staining patterns also shown are diffuse nuclear staining or no BrdU incorporation.

a consequence of the introduced p53 alleles. Further, these results indicate that the increase in APB-positive cells does not require the transactivation activity of p53. It has previously been demonstrated that APBs are cell cycle regulated, accumulating in cells in the late S and G₂/M phases of the cell cycle (22, 62). The increase in APB-positive cells at 32°C suggests that the expression of ts-p53TA may perturb the cell cycle in cells that use ALT for telomere maintenance.

p53 causes an S phase delay in cells that use ALT for telomere maintenance. To determine if the expression of p53-TA altered cell cycle progression in ALT cells relative to telomerase-positive cells, asynchronous cultures of the HIO107-TA and the HIO114-TA cell lines were pulsed with BrdU after 72 h of growth at either 32 or 39°C, followed by simultaneous detection of BrdU and DNA content and FACS analysis. Both the HIO107-TA ALT and the HIO114-TA telomerase-positive cell lines incorporated BrdU throughout S phase at 39°C (Fig. 3B), consistent with their normal growth at this temperature (Fig. 2C). The presence of active p53-TA, i.e., when cells are grown at 32°C, had no effect on BrdU incorporation in the HIO114-TA cell line, which is consistent with the continued incorporation in transiently transfected populations (Fig. 1) and with the continued proliferation at 32°C in clonal cell lines that stably express this allele (Fig. 2C). In contrast, the BrdU-positive cells in the HIO107-TA cell line at 32°C are concentrated in early S phase (Fig. 3B), suggesting that p53 causes a transactivation-independent delay of normal S phase progression specifically in ALT cells.

To further investigate this ALT-specific effect on S phase, the HIO107-TA cell line was arrested in early S phase with HU and then cultured at either 32 or 39°C following the removal of drug. Because the expression of ts-p53TA does not perturb cell cycle progression in the telomerase-positive HIO114 cells (Fig. 3B), they were not assessed here. The HIO107 parental cell line progresses through the cell cycle at both temperatures following the removal of HU (Fig. 3C). Likewise, HIO107-TA cells arrest efficiently and at 39°C, when ts-p53TA is inactive, progress through the cell cycle normally after the removal of HU. In contrast, HIO107-TA cells do not progress through the cell cycle at 32°C following the removal of HU and instead appear to be trapped in early S phase (Fig. 3C). A similar S phase delay is observed in the HIO107-WT cell line (data not shown). Quantitation of the FACS data indicates that HIO107-TA cells cultured at 32°C continue to accumulate in S phase for at least 24 h following the removal of HU (Fig. 3D). BrdU labeling for 40 min prior to harvest indicates that HIO107-TA cells are replicating DNA at 32°C at every time point, albeit to a much reduced extent (data not shown). Thus, the HIO107-TA cells are not arrested in S phase but instead exhibit a delayed progression with reduced DNA replication as measured by BrdU incorporation. These data confirm that p53 causes a transactivation-independent S phase delay in ALT cells.

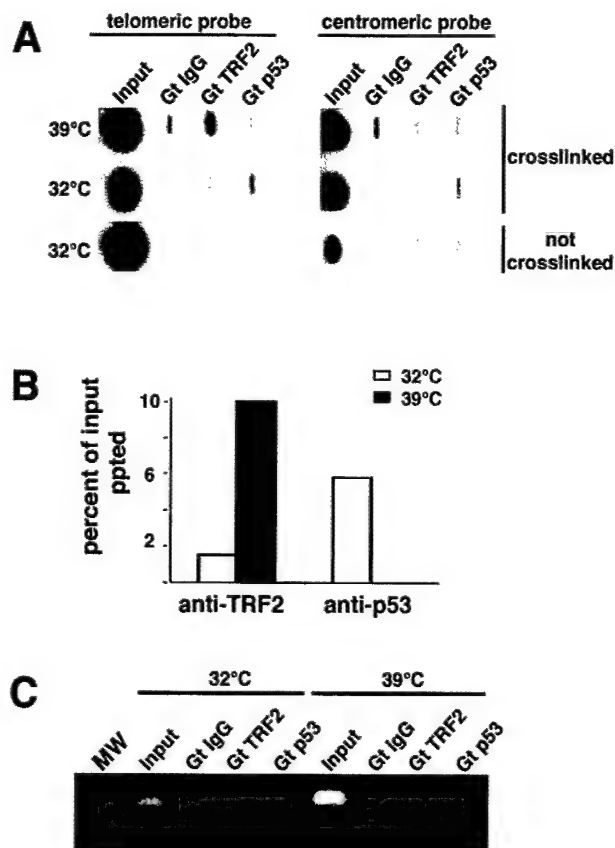
To directly visualize the DNA replication that occurs during the S phase delay in HIO107-TA cells, cultures were grown for 48 h at either 39 or 32°C and then for an additional 24 h in the presence of BrdU. The frequency of cells that incorporated BrdU was scored in at least 100 cells for each culture. At 39°C, >90% of the cells incorporated high levels of BrdU, defined as total nuclear staining (Fig. 3E). In contrast, at 32°C approxi-

mately 65% of the HIO107-TA cells incorporated high levels of BrdU (Fig. 3E). A large fraction of cells at 32°C, approximately 30%, did not incorporate any BrdU. In addition, in approximately 5% of the cells cultured at 32°C, BrdU incorporation was punctate with the BrdU colocalizing with TRF2 and TRF1 at APBs (Fig. 3E and F and data not shown). This pattern was extremely rare at 39°C, occurring in <1% of the cells. BrdU incorporation at APBs has been noted previously and has been suggested to represent replication of ALT telomeres (61). These data suggest that some of the continued BrdU incorporation observed in the S phase-delayed populations is occurring at, or near, telomeric DNA.

p53 associates with the telomeric complex in cells that use ALT for telomere maintenance. It has been reported that p53 binds to telomeric DNA in vitro (51). If the ts-p53TA allele used here is exerting its growth inhibitory effect by preventing telomeric recombination, then we might predict an association between p53 and the telomeric complex. To determine if p53 is able to directly interact with telomeres, we utilized ChIP assays. As a positive control, ChIP assays were carried out by using the constitutive telomere binding protein TRF2 (6) (Fig. 4A). Immunoprecipitation is dependent upon TRF2 antibody and requires protein-DNA cross-linking with formaldehyde. Similar results were obtained with an antibody against the telomeric binding protein TRF1 (data not shown). Furthermore, immunoprecipitation is specific for telomeric DNA because there is no detectable hybridization when the same filter is hybridized with probes directed against centromeric α -satellite sequences (Fig. 4A). We found that p53 is associated with the telomeric complex in the HIO107-TA cell line but only when p53 is in an active conformation, i.e., when the cells are cultured at 32°C (Fig. 4A). The enrichment in telomeric DNA precipitated at 32°C by antibodies against p53 is paralleled by a similar level of enrichment in α -satellite sequences (4.6% of input), suggesting that when p53 is competent to bind to DNA, it is located at numerous sites in the genome. Quantitation of the amount of telomeric DNA precipitated is shown in Fig. 4B.

To confirm that the antibody against p53 was precipitating known p53-associated DNA sequences, a proportion of the immunoprecipitated DNA was interrogated by PCR for the presence of the p21^{WAF1} promoter. As expected, the region of the p21^{WAF1} promoter encompassing the p53 binding site at position -1.4 kb (18) was precipitated by the antibody against p53 in both cell lines at 32°C (Fig. 4D), but not by nonspecific rabbit IgG, goat IgG, or anti-TRF2.

Growth suppression of ALT cells by p53 requires intact suppression of recombination function. p53 is able to bind to Holliday junctions (29) in vitro and, through interactions with the BLM and WRN proteins, may contribute to the suppression of inappropriate recombination (64). To directly test if the transactivation-independent, ALT-specific growth inhibition is due to interference with recombination-based telomere maintenance, we introduced into the ts-p53TA plasmid the R175H, R248Q, and R273H mutations, which compromise the specific DNA binding and suppression of recombination functions of p53 (3, 13, 15, 45, 60). While some p53 mutations have been reported to have gain-of-function activities, a functional transactivation domain appears to be essential for this phenotype (9, 49) and, thus, would be eliminated in the ts-p53TA mutant background. Because the mutations that impair suppression of



the recombination function fall within the DNA binding domain of p53, these mutants are unable to associate with p53 target sequences (13, 60). HIO107 ALT and HIO114 telomerase-positive cells were transiently transfected with these compound mutant alleles and analyzed as shown in Fig. 1. Again, only those cells with high levels of p53, therefore expressing the ectopic allele, were scored (Fig. 5A, top). Furthermore, Western blot analysis indicated that all mutant proteins are stable (Fig. 5A, bottom). As shown previously, the expression of ts-p53TA inhibits BrdU incorporation at 32°C in the HIO107 and WI38-VA13/2RA ALT cell lines but not in the telomerase-positive HIO114 cell line. The introduction of any of the com-

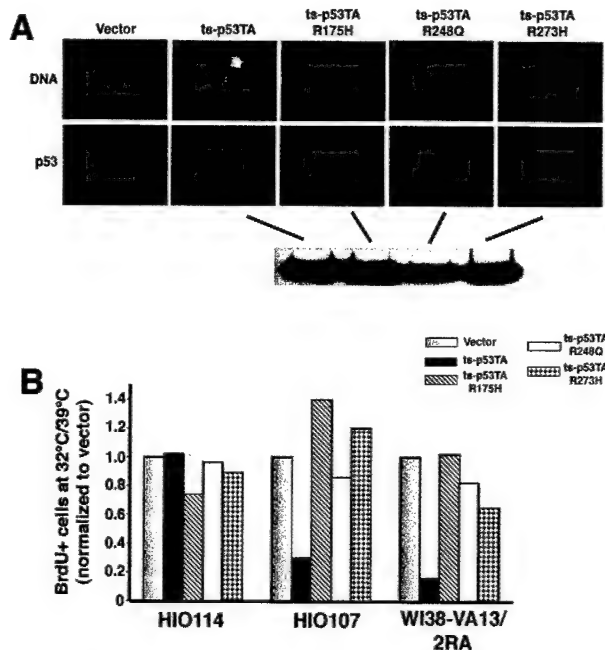


FIG. 5. Expression of p53 mutant alleles that compromise the ability of p53 to suppress recombination does not inhibit DNA replication in ALT cell lines. (A) Immunofluorescent staining of the indicated p53 compound mutant protein (top). DNA was stained with DAPI. Western analysis of each p53 compound mutant allele is shown (bottom). (B) Ratio of BrdU incorporation in transfected cells at 39 versus 32°C for the HIO107 and WI38-VA13/2RA ALT-positive and HIO114 telomerase-positive cell lines transfected with empty vector or the indicated p53 allele. All ratios were normalized to the frequency of BrdU incorporation at the two temperatures in control cells transfected with empty vector in parallel.

pound mutants had no effect in the telomerase-positive HIO114 cell line, with BrdU incorporation occurring at a level similar to that seen with the ts-p53TA allele. However, the expression of the ts-p53TAR175H, ts-p53TA-R248Q, or ts-p53TA-R273H allele in either the HIO107 or the WI38-VA13/2RA ALT cell lines suppressed the inhibition of BrdU incorporation (Fig. 5B). These data indicate that the transactivation-independent, ALT-specific inhibition of DNA replication caused by p53 requires the specific DNA binding and suppression of the recombination functions of p53.

DISCUSSION

Present evidence suggests that ALT is a recombination-based mechanism (17) and that the mutation of p53 provides a permissive environment for the activation of this telomere maintenance pathway (8, 39, 50). Because p53 inhibits homologous recombination (3, 34, 58), we hypothesized that p53 might regulate the activation of ALT via its role in regulating inappropriate recombination. Here we demonstrate that transactivation-incompetent p53 causes growth inhibition in ALT but not in telomerase-positive cells. Growth inhibition is associated with a specific early S phase delay in ALT cells. In addition, we demonstrate that p53 may associate with telomeric DNA in vivo. As expected for a DNA binding protein, p53 is also associated with other sequences in the genome, such

as α -satellite DNA and the p21^{WAF1} promoter. However, our data indicate that p53 may also associate with the telomere, making it possible that the growth inhibition seen in ALT cells, at least in part, is a result of p53 acting directly at the telomere. Finally, we show that p53 alleles defective in specific DNA binding and recombination suppression no longer inhibit DNA replication in ALT cells. These data are consistent with a model whereby p53 differentially inhibits growth of ALT cells by perturbing recombination between telomeres in a mechanism that requires a direct interaction of p53 with the telomeric complex. Alternatively, the processes underlying ALT might be detrimental to cellular survival, for example, by being associated with high levels of genome instability. In this scenario, the growth inhibition of ALT cells would not require either a direct inhibition of telomere maintenance by p53 or direct activation by dysfunctional telomeres of a p53-dependent pathway. Instead, elimination of this pathway would be necessary to allow cells to circumvent constitutive DNA damage checkpoint signaling. Because the p53 alleles used here have no transactivation activity, growth arrest induced by such a constitutive DNA damage signal would be elicited via some other activity of p53, such as transcriptional repression (46).

It has been demonstrated that p53 associates with the T loop *in vitro* (51), and thus p53 may be an integral component of the telomere, normally acting to prevent extension of telomere sequences from this partial Holliday junction. Regardless of whether p53 associates specifically with telomeres in ALT cells or is a constitutive component of all telomeres, in the model proposed above it would only act to limit growth in cells that are dependent upon recombination to maintain telomeres. Less telomeric DNA is reproducibly precipitated by anti-TRF2 from cells cultured at 32 than at 39°C. Although TRF2 is present on telomeres throughout the cell cycle (6), it is possible that there is a transient dissociation of TRF2 from the telomere when telomeres are elongated by ALT, which might become detectable due to the p53-induced enrichment of cells in S phase. In light of the potentially dynamic association of TRF2 with telomeres, it is intriguing to note the recent reports that dysfunctional telomeres actively engage other components of the DNA damage checkpoint machinery (14, 54). The direct demonstration of p53 association with telomeres in primary cells or telomerase-positive cells *in vivo* and the conditions under which such an association might occur have yet to be established and are topics worthy of further study.

All the ALT cell lines used here contain SV40 large T-Ag, which itself has been shown to have recombination promoting activity (52). Because ALT is a recombination-based mechanism, the presence of large T-Ag might affect this pathway. However, because both the telomerase-positive HIO114 and Meso6 cell lines analyzed here also contain SV40 (1, 19), any differences observed between the telomerase-positive and ALT cell lines cannot be attributable simply to the presence of large T-Ag. In addition, the results obtained here might have been confounded by the presence of endogenous p53, which in the case of the HIO107 and HIO114 cell lines has been determined to remain wild type (A.K. Godwin, unpublished results), in that overexpression of the exogenous p53 alleles might result in the release of endogenous p53 from large T-Ag. However, it is apparent that very little, if any, endogenous p53 is active in the cell lines expressing the p53TA allele because the p53-

responsive p21^{WAF1} protein does not accumulate even after 48 h. Furthermore, the exogenous p53WT allele does not induce a phenotype at 39°C, indicating that the inactive protein does not compete with the endogenous p53 for binding to large T-Ag. This is consistent with observations that many mutant p53 proteins do not bind to large T-Ag (66). Thus, the phenotypes observed are most likely the result of the expression of the exogenous p53 alleles rather than a nonspecific effect of large T-Ag or the result of the endogenous p53 protein. Because these experiments were carried out using overexpressed protein, it is possible that endogenous p53 levels may not accumulate to sufficient levels to generate the growth inhibition phenotype. However, this is unlikely, given that the stabilization and accumulation of protein levels in response to DNA damage are a central regulatory mechanism for p53 (9).

Willers and coworkers (59) investigated various alleles of p53 with respect to their ability to suppress homologous recombination in mice. The homologous murine ts-p53 allele, V135A, exhibited a severely impaired DNA damage response at 37°C, indicating that a large proportion of the protein is in mutant conformation at this temperature. While the ts-p53 allele was still able to suppress recombination at 37°C, the activity was impaired relative to wild-type protein. Because these experiments were carried out at 37°C, the protein may retain some wild-type activity. In contrast to the study cited above, our analysis was carried out with human cells. In addition, the cultures used here were maintained at 39°C to ensure that the protein was in mutant conformation and unable to bind to DNA. Finally, p53 recognizes and binds to three-stranded DNA substrates *in vitro*, and its ability to suppress recombination *in vivo* requires both an intact core domain and an oligomerization domain (15). Thus, specific DNA binding by p53 is essential for the suppression of recombination. This is consistent with our ability to isolate clonal cell lines expressing ts-p53 alleles at 39°C when DNA binding and the suppression of recombination function are impaired (59). Taken together, the data indicate that ts-p53 is impaired in its ability to suppress recombination.

The suppression of recombination by p53 may occur through the recognition of altered DNA structures, such as single-stranded regions or cruciforms, and the targeting of cells containing these structures for apoptosis. However, the suppression of recombination by p53 does not require the transactivation function (15, 59), suggesting that p53 exerts its effect on recombination independent of gene expression. p53 is stabilized in S phase in response to replication blocks (21) or hypoxia (24). Although the stabilized protein is incompetent to induce many target genes (21), these observations suggest that p53 may have transactivation-independent functions in S phase, which in turn may be responsible for the delay of the normal S phase progression observed here. We propose one model in which the growth inhibition observed in the HIO107-TA cell lines is due to the perturbation of ALT-dependent telomere maintenance by p53. Although ALT cells contain many extremely long telomeres, growth inhibition occurs rapidly, within 48 h of ts-p53TA expression. We were unable to detect a global, catastrophic loss of telomeric sequences to account for this rapid induction of growth inhibition. This may be because ALT cells also contain a few chromosomes with extremely short telomeres (10, 40) that may

activate telomere length-based checkpoints (14, 54) if not elongated. An analysis of telomeric end protection function, assayed by quantitating the frequency of anaphase bridges in mitosis, is precluded due to the growth inhibition occurring in S phase. In addition, because the growth defect seen here is different from the p53-dependent apoptotic response observed when telomere dysfunction is forced through the expression of dominant-negative TRF2 (28), the S phase delay is likely activated through a different pathway. Reintroduction of telomerase into ALT cell lines causes the preferential elongation of the short telomeres present in ALT cells (10, 23, 40) but, except in rare cases (20), does not result in the suppression of ALT (10, 23, 40). These cell lines provide a means of differentiating which of these two signals, critically short telomeres versus illegitimate recombination intermediates, is actually responsible for the observed growth inhibition. The identification of the initiating signal and downstream pathway generating the S phase delay seen here is worthy of future study.

We and others have previously demonstrated that APB-positive cells accumulate in the late S and G₂/M phases of the cell cycle (22, 62) and suggested that these structures are coordinately regulated with the cell cycle. However, in the experiments conducted here, the presence of p53 causes both an accumulation of APB-positive cells and an accumulation of cells in early S phase. These results suggest that the monitoring of telomeres and/or the recombination reaction underlying telomere extension in ALT cells occurs in early S phase. The difference in the timing of the accumulation of APB-positive cells may be due to the absence of p53 in the cell lines previously studied, resulting in a delayed cellular response. In this scenario, APBs would be sites of telomeric recombination, which are stalled early in S phase by p53. BrdU incorporation at APBs has been reported previously and suggested to be a consequence of telomere replication by the ALT pathway (61). If so, then the BrdU incorporation at foci observed in the HIO107-TA cells at 32°C may consist of cells that have overcome the p53-dependent stall in S phase. Alternatively, APBs may form as a downstream response to a checkpoint signal, the timing of which is p53 dependent.

While loss of p53 has been correlated with a predilection for the activation of ALT rather than telomerase for telomere maintenance, the experiments described here are the first to directly test the role of p53 in ALT. Our data support a model whereby p53 plays a role in the regulation of ALT by perturbing telomeric recombination. Although this model provides one regulatory mechanism for the activation of ALT, it may not be the only one. ALT tumors derived from telomerase-deficient mice retain active p53, as judged by a robust DNA damage response (11). However, because the DNA damage response and suppression of recombination functions of p53 are separable (15, 59), it is not clear if these tumors are compromised for recombination suppression. Furthermore, tumors arising in mice deficient in both poly(ADP-ribose) polymerase and p53 undergo telomere maintenance by ALT (56). This only occurs in the absence of p53, suggesting that p53 may also suppress activation of ALT in mice under the appropriate conditions. The data presented here suggest that p53-compromised cells may be more likely to activate ALT in response to telomerase inhibition than cells containing wild-type p53,

clearly an important issue with respect to the therapeutic value of telomerase inhibitors in the clinic.

ACKNOWLEDGMENTS

This work was supported by N.I.H. grants CA098087-01 (to D.B.), CA-45745-14 (to J.R.T.), and CA006927 (to Fox Chase Cancer Center) and D.O.D grant DAMD17-01-1-0724 (to D.B.).

We thank P. Adams and X. Ye for advice with cell cycle analysis and P. Adams, J. Chernoff, D. Kipling, and K. Zaret for critical reading of the manuscript.

REFERENCES

1. Auersperg, N., S. Maines-Bandiera, J. Booth, H. Lynch, A. Godwin, and T. Hamilton. 1995. Expression of two mucin antigens in cultured human ovarian surface epithelium: Influence of a family history of ovarian cancer. *Am. J. Obstet. Gynecol.* 173:558-565.
2. Bayne, R. A., D. Broccoli, M. H. Taggart, E. J. Thomson, C. J. Farr, and H. J. Cooke. 1994. Sandwiching of a gene within 12 kb of a functional telomere and alpha satellite does not result in silencing. *Hum. Mol. Genet.* 3:539-546.
3. Bertrand, P., D. Rouillard, A. Boulet, C. Levalois, T. Soussi, and B. S. Lopez. 1997. Increase of spontaneous intrachromosomal homologous recombination in mammalian cells expressing a mutant p53 protein. *Oncogene* 14: 1117-1122.
4. Boehden, G. S., N. Akyuz, K. Roemer, and L. Wiesmuller. 2003. p53 mutated in the transactivation domain retains regulatory functions in homology-directed double-strand break repair. *Oncogene* 22:4111-4117.
5. Broccoli, D., L. A. Godley, L. A. Donehower, H. E. Varmus, and T. de Lange. 1996. Telomerase activation in mouse mammary tumors: lack of detectable telomere shortening and evidence for regulation of telomerase RNA with cell proliferation. *Mol. Cell. Biol.* 16:3765-3772.
6. Broccoli, D., A. Smogorzewska, L. Chong, and T. de Lange. 1997. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat. Genet.* 17:231-235.
7. Bryan, T. M., A. Englezou, L. Dalla-Pozza, M. A. Dunham, and R. R. Reddel. 1997. Evidence for an alternative mechanism for maintaining telomere length in human tumors and tumor-derived cell lines. *Nat. Med.* 3:1271-1274.
8. Bryan, T. M., A. Englezou, J. Gupta, S. Bacchetti, and R. R. Reddel. 1995. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* 14:4240-4248.
9. Cadwell, C., and G. P. Zambetti. 2001. The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. *Gene* 277:15-30.
10. Cerone, M. A., J. A. Londono-Vallejo, and S. Bacchetti. 2001. Telomere maintenance by telomerase and by recombination can coexist in human cells. *Hum. Mol. Genet.* 10:1945-1952.
11. Chang, S., C. M. Khoo, M. L. Naylor, R. S. Maser, and R. A. DePinho. 2003. Telomere-based crisis: functional differences between telomerase activation and ALT in tumor progression. *Genes Dev.* 17:88-100.
12. Chen, Q., A. Ijima, and C. W. Greider. 2001. Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. *Mol. Cell. Biol.* 21:1819-1827.
13. Cho, Y., S. Gorina, P. D. Jeffrey, and N. P. Pavletich. 1994. Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations. *Science* 265:346-355.
14. d'Adda di Fagagna, F., P. M. Reaper, L. Clay-Farrace, H. Fiegler, P. Carr, T. Von Zglinicki, G. Saretzki, N. P. Carter, and S. P. Jackson. 2003. A DNA damage checkpoint response in telomere-initiated senescence. *Nature* 426: 194-198.
15. Dudenhofer, C., M. Kurth, F. Janus, W. Deppert, and L. Wiesmuller. 1999. Dissociation of the recombination control and the sequence specific trans-activation function of p53. *Oncogene* 18:5773-5784.
16. Dudenhofer, C., G. Rohaly, K. Will, W. Deppert, and L. Wiesmuller. 1998. Specific mismatch recognition in heteroduplex intermediates by p53 suggests a role in fidelity control of homologous recombination. *Mol. Cell. Biol.* 18:5332-5342.
17. Dunham, M. A., A. A. Neymann, C. L. Fasching, and R. R. Reddel. 2000. Telomere maintenance by recombination in human cells. *Nat. Genet.* 26: 447-450.
18. el-Deiry, W. S., T. Tokino, V. E. Velculescu, D. B. Levy, R. Parsons, J. M. Trent, D. Lin, W. E. Mercer, K. W. Kinzler, and B. Vogelstein. 1993. WAF1, a potential mediator of p53 tumor suppression. *Cell* 75:817-825.
19. Foddis, R., A. De Rienzo, D. Broccoli, M. Bocchetta, E. Stekala, P. Rizzo, A. Tosolini, J. V. Grobely, S. C. Jhanwar, H. I. Pass, J. R. Testa, and M. Carbone. 2002. SV40 infection induces telomerase activity in human mesothelial cells. *Oncogene* 21:1434-1442.
20. Ford, L. P., Y. Zou, K. Pongracz, S. M. Gryaznov, J. W. Shay, and W. E. Wright. 2001. Telomerase can inhibit the recombination-based pathway of telomere maintenance in human cells. *J. Biol. Chem.* 276:32198-32203.

21. Gottifredi, V., S. Shieh, Y. Taya, and C. Prives. 2001. p53 accumulates but is functionally impaired when DNA synthesis is blocked. *Proc. Natl. Acad. Sci. USA* 98:1036-1041.
22. Grobely, J. V., A. K. Godwin, and D. Broccoli. 2000. ALT-associated PML bodies are present in viable cells and are enriched in cells in the G₂/M phase of the cell cycle. *J. Cell Sci.* 113:4577-4585.
23. Grobely, J. V., M. Kulp-McEliece, and D. Broccoli. 2001. Effects of reconstitution of telomerase activity on telomere maintenance by the alternative lengthening to telomeres (ALT) pathway. *Hum. Mol. Genet.* 10:1953-1961.
24. Hammond, E. M., N. C. Denko, M. J. Dorie, R. T. Abraham, and A. J. Giaccia. 2002. Hypoxia links ATR and p53 through replication arrest. *Mol. Cell. Biol.* 22:1834-1843.
25. Hemann, M. T., M. A. Strong, L.-Y. Hao, and C. W. Greider. 2001. The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability. *Cell* 107:67-77.
26. Henson, J. D., A. A. Neumann, T. R. Yeager, and R. R. Reddel. 2002. Alternative lengthening of telomeres in mammalian cells. *Oncogene* 21:598-610.
27. Jimenez, G. S., M. Nister, J. M. Stommel, M. Beeche, E. A. Barcarse, X. Q. Zhang, S. O'Gorman, and G. M. Wahl. 2000. A transactivation-deficient mouse model provides insights into Trp53 regulation and function. *Nat. Genet.* 26:37-43.
28. Karlseider, J., D. Broccoli, Y. Dai, S. Hardy, and T. de Lange. 1999. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* 283:1321-1325.
29. Lee, S., L. Cavallo, and J. Griffith. 1997. Human p53 binds Holliday junctions strongly and facilitates their cleavage. *J. Biol. Chem.* 272:7532-7539.
30. Lee, S., B. Elenbaas, A. Levine, and J. Griffith. 1995. p53 and its 14 kDa C-terminal domain recognize primary DNA damage in the form of insertion/deletion mismatches. *Cell* 81:1013-1020.
31. Lin, J., J. Chen, B. Elenbaas, and A. Levine. 1994. Several hydrophobic amino acids in the p53 amino-terminal domain are required for transcriptional activation, binding to mdm-2 and the adenovirus 5 E1B 55-kD protein. *Genes Dev.* 8:1235-1246.
32. Lundblad, V., and E. H. Blackburn. 1993. An alternative pathway for yeast telomere maintenance rescues est1-senescence. *Cell* 73:347-360.
33. Martinez, J., I. Georgoff, J. Martinez, and A. J. Levine. 1991. Cellular localization and cell cycle regulation by a temperature-sensitive p53 protein. *Genes Dev.* 5:151-159.
34. Mekeel, K. L., W. Tang, L. A. Kachnic, C. M. Luo, J. S. DeFrank, and S. N. Powell. 1997. Inactivation of p53 results in high rates of homologous recombination. *Oncogene* 14:1847-1857.
35. Michalovitz, D., D. Halevy, and M. Oren. 1990. Conditional inhibition of transformation and of cell proliferation by a temperature-sensitive mutant of p53. *Cell* 62:671-680.
36. Miller, D. A., V. Sharma, and A. R. Mitchell. 1988. A human-derived probe, p82H, hybridizes to the centromeres of gorilla, chimpanzee, and orangutan. *Chromosoma* 96:270-274.
37. Nelson, D. M., X. Ye, C. Hall, H. Santos, T. Ma, G. D. Kao, T. J. Yen, J. W. Harper, and P. D. Adams. 2002. Coupling of DNA synthesis and histone synthesis in S phase independent of cyclin/cdk2 activity. *Mol. Cell. Biol.* 22:7459-7472.
38. Niida, H., Y. Shinkai, M. P. Hande, T. Matsumoto, S. Takehara, M. Tachibana, M. Oshimura, P. M. Lansdorp, and Y. Furuchi. 2000. Telomere maintenance in telomerase-deficient mouse embryonic stem cells: characterization of an amplified telomeric DNA. *Mol. Cell. Biol.* 20:4114-4127.
39. Opitz, O. G., Y. Suliman, W. M. Hahn, H. Harada, H. E. Blum, and A. K. Rustig. 2001. Cyclin D1 overexpression and p53 inactivation immortalize primary oral keratinocytes by a telomerase-independent mechanism. *J. Clin. Invest.* 108:725-732.
40. Perrem, K., L. M. Colgin, A. A. Neumann, T. R. Yeager, and R. R. Reddel. 2001. Coexistence of alternative lengthening of telomeres and telomerase in hTERT-transfected GM847 cells. *Mol. Cell. Biol.* 21:3862-3875.
41. Quignon, F., F. DeBels, M. Koken, J. Feunteun, J. C. Ameisen, and H. de The. 1998. PML induces a novel caspase-independent death process. *Nat. Genet.* 20:259-265.
42. Reddel, R. R., T. M. Bryan, and J. P. Murnane. 1997. Immortalized cells with no detectable telomerase activity. A review. *Biochemistry (Moscow)* 62:1254-1262.
43. Roemer, K., and N. Mueller-Lantzsch. 1996. p53 transactivation domain mutant Q22, S23 is impaired for repression of promoters and mediation of apoptosis. *Oncogene* 12:2069-2079.
44. Sabbatini, P., J. Lin, A. J. Levine, and E. White. 1995. Essential role for p53-mediated transcription in E1A-induced apoptosis. *Genes Dev.* 9:2184-2192.
45. Saintigny, Y., D. Rouillard, B. Chaput, T. Soussi, and B. S. Lopez. 1999. Mutant p53 proteins stimulate spontaneous and radiation-induced intra-chromosomal homologous recombination independently of the alteration of the transactivation activity and of the G1 checkpoint. *Oncogene* 18:3553-3563.
46. Sang, B. C., J. Y. Chen, J. Minna, and M. S. Barbosa. 1994. Distinct regions of p53 have a differential role in transcriptional activation and repression functions. *Oncogene* 9:853-859.
47. Seeler, J. S., and A. Dejean. 1999. The PML nuclear bodies: actors or extras? *Curr. Opin. Genet. Dev.* 9:362-367.
48. Shay, J. W., and S. Bacchetti. 1997. A survey of telomerase activity in human cancer. *Eur. J. Cancer* 33:787-791.
49. Sigal, A., and V. Rotter. 2000. Oncogenic mutations of the p53 tumor suppressor: the demons of the guardian of the genome. *Cancer Res.* 60:6788-6793.
50. Sood, A. K., J. Coffin, J. Sarvenaz, R. E. Buller, M. J. C. Hendrix, and A. Klingelutz. 2002. p53 null mutations are associated with a telomerase negative phenotype in ovarian carcinoma. *Cancer Biol. Ther.* 1:511-517.
51. Stansel, R. M., D. Subramanian, and J. D. Griffith. 2002. p53 binds telomeric single strand overhangs and t-loop junctions in vitro. *J. Biol. Chem.* 277:11625-11628.
52. St-Onge, L., L. Bouchard, and M. Bastin. 1993. High-frequency recombination mediated by polyomavirus large T antigen defective in replication. *J. Virol.* 67:1788-1795.
53. Sturzbecher, H. W., B. Donzelmann, W. Henning, U. Knippschild, and S. Buchhop. 1996. p53 is linked directly to homologous recombination processes via RAD51/RecA protein interaction. *EMBO J.* 15:1992-2002.
54. Takai, H., A. Smogorzewska, and T. de Lange. 2003. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* 13:1549-1556.
55. Teng, S.-C., and V. A. Zakian. 1999. Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 19:8083-8093.
56. Tong, W. M., M. P. Hande, P. M. Lansdorp, and Z. Q. Wang. 2001. DNA strand break-sensing molecule poly(ADP-ribose) polymerase cooperates with p53 in telomere function, chromosome stability, and tumor suppression. *Mol. Cell. Biol.* 21:4046-4054.
57. Wiesmuller, L., J. Cammenga, and W. W. Deppert. 1996. In vivo assay of p53 function in homologous recombination between simian virus 40 chromosomes. *J. Virol.* 70:737-744.
58. Willers, H., E. E. McCarthy, W. Alberti, J. Dahm-Daphi, and S. N. Powell. 2000. Loss of wild-type p53 function is responsible for upregulated homologous recombination in immortal rodent fibroblasts. *Int. J. Radiat. Biol.* 76:1055-1062.
59. Willers, H., E. E. McCarthy, B. Wu, H. Wunsch, W. Tang, D. G. Taghian, F. Xia, and S. N. Powell. 2000. Dissociation of p53-mediated suppression of homologous recombination from G1/S cell cycle checkpoint control. *Oncogene* 19:632-639.
60. Wong, K. B., B. S. DeDecker, S. M. Freund, M. R. Proctor, M. Bycroft, and A. R. Fersht. 1999. Hot-spot mutants of p53 core domain evince characteristic local structural changes. *Proc. Natl. Acad. Sci. USA* 96:8438-8442.
61. Wu, G., X. Jiang, W. H. Lee, and P. L. Chen. 2003. Assembly of functional ALT-associated promyelocytic leukemia bodies requires Nijmegen breakage syndrome 1. *Cancer Res.* 63:2589-2595.
62. Wu, G., W. H. Lee, and P. L. Chen. 2000. NBS1 and TRF1 colocalize at promyelocytic leukemia bodies during late S/G2 phases in immortalized telomerase-negative cells. Implication of NBS1 in alternative lengthening of telomeres. *J. Biol. Chem.* 275:30618-30622.
63. Wu, L., J. H. Bayle, B. Elenbaas, N. P. Pavletich, and A. J. Levine. 1995. Alternatively spliced forms in the carboxy-terminal domain of the p53 protein regulate its ability to promote annealing of complementary single strands of nucleic acids. *Mol. Cell. Biol.* 15:497-504.
64. Yang, Q., R. Zhang, X. W. Wang, E. A. Spillare, S. P. Linke, D. Subramanian, J. D. Griffith, J. L. Li, I. D. Hickson, J. C. Shen, L. A. Loeb, S. J. Mazur, E. Appella, R. M. Brosh, Jr., P. Karmakar, V. A. Bohr, and C. C. Harris. 2002. The processing of Holliday junctions by BLM and WRN helicases is regulated by p53. *J. Biol. Chem.* 277:31980-31987.
65. Yeager, T., A. Neumann, A. Englezou, L. Huschtscha, J. Noble, and R. Reddel. 1999. Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Res.* 59:4175-4179.
66. Zambetti, G. P., and A. J. Levine. 1993. A comparison of the biological activities of wild-type and mutant p53. *FASEB J.* 7:855-865.

Microarrays in Cancer: Research and Applications

Michael F. Ochs¹ and Andrew K. Godwin²

¹Bioinformatics, Department of Information Science and Technology and ²Department of Medical Oncology, Fox Chase Cancer Center, Philadelphia, PA, USA

BioTechniques 34:S4-S15 (March 2003)

INTRODUCTION

Cancer is an important public health concern in the United States and around the world. After heart disease, it is the second leading cause of death, accounting for 23.0% of deaths in the U.S. (1). In the year 2003, it is estimated that about 1 334 100 new cases of invasive cancer will be diagnosed in the U.S. (2) and that this number will double by the year 2050 (3). In addition, over one million cases of basal and squamous cell skin cancer will be diagnosed annually (1,2). It is estimated that 556 500 Americans will die from cancer in 2003, corresponding to 1500 deaths per day. Among men, the most common cancers are cancers of the prostate (33%), lung and bronchus (14%), and colon and rectum (11%) (2). Among women, the three most commonly diagnosed cancers are cancers of the breast (32%), lung and bronchus (12%), and colon and rectum (11%) (2). In 1987, lung cancer surpassed breast cancer as the leading cause of cancer death in women and is expected to account for at least 25% of all female cancer deaths in the coming years. The good news is that following more than 70 years of increases, the recorded number of total cancer deaths among men and women in the U.S. has declined for the first time. For all cancer types and races combined, cancer death rates declined by 1.5% per year in males and by 0.6% per year in females from 1992 to 1999. These somewhat encouraging trends are primarily associated with improved screening techniques and the subsequent increase in diagnosis at an early stage when most, but not all cancers are more successfully treated. Unfortunately, most current cancer therapies have limited efficacy in curing late-stage disease. Therefore, there continues to be a need to develop new approaches to (i) diagnose cancer early in its clinical course, (ii) more effectively treat advanced stage disease, (iii) better predict a tumor's response to therapy prior to the actual treatment, and (iv) ultimately prevent disease from arising through chemopreventive strategies. These goals can only be accomplished through a better understanding of how certain genes and their encoded proteins contribute to disease onset and tumor progression and how they influence the response of patients to drug therapies. Innovations in genetic, biological, and biochemical approaches are necessary to realize these goals.

In this aspect, we are in an era of unprecedented opportunity for addressing these questions and uncovering the molecular basis of cancer. It is now clear that new approaches to science,

based on comprehensive molecular analysis, offer remarkable new opportunities to enhance the framework of our knowledge of cancer. It has been evident from the beginning of the molecular study of cancer that this is not one disease but many, and that any particular cancer arises as the result of the gradual accumulation of genetic changes within a single cell. New technologies offer the potential to describe specific types of genetic changes and the patterns of altered gene product expression and function that define the actual definition of any cancer in the context of, but not strictly dependent upon, its site of origin. We anticipate that these new technologies will lead to the identification of previously undetectable features of the molecular basis of individual tumor characteristics and profile progression and response to therapy. What is particularly evident is the need to link the information gained from these new technologies with a wealth of clinical information (obtained from pathologists to oncologists) associated with each clinical sample. DNA microarrays are one of these new powerful technologies. In the 8 years since its inception, microarray technology has become a major tool for the investigation of global gene expression of all aspects of human disease and in biomedical research. The range of applications of microarray technology is enormous. Recent studies in human cancer have demonstrated that microarrays can be used to develop a new taxonomy of cancer, including major insights into the genesis, progression, prognosis, and response to therapy on the basis of gene expression profiles. We are just beginning to be able to answer clinically important questions about cancer, such as which tumors will behave aggressively, which will remain indolent, and which are likely to respond to different therapies. Furthermore, the list of potential uses of this technique is not limited to cancer research as evident by the exponentially growing number of manuscripts that have recently been published using these methods (Figure 1).

CANCER AND LARGE-SCALE GENE EXPRESSION CHANGES

The concept of "discovery science" is best illustrated by the human genome project (4,5). This type of approach involves the identification of the components of a system without the prior formulation of hypotheses as to how these components function (6). This scientific method has spawned many global genomic

approaches to begin characterizing a vast number of biological systems. Traditionally, studies measuring the differences between cell types or cell pathways following perturbation of the cell have been carried out at the level of transcribed mRNA using methods such as differential display (DD) (7,8), representational difference analysis (RDA) (9), and suppressive subtractive hybridization (SSH) (10,11). However, these methods, although fruitful and still in use, have limited scope in terms of the number of gene expression patterns that can be analyzed. Therefore, more robust methods, including serial analysis of gene expression (SAGE) (12), massive parallel signature sequencing (MPSS) (13), and cDNA arrays and chip technologies (14–18) have taken their places.

Development of Microarrays

DNA microarrays are rapidly becoming a fundamental tool in genomic research. Microarray methods were initially developed to study differential gene expression using complex populations of RNA (14–18). Refinement of these methods now permits the analysis of copy number imbalances and gene amplification of DNA (19) and the detection of small deletions or insertions in tumor suppressor genes (20), and they have recently been applied to the systematic analysis of expression at the protein level (21). For the purpose of this review we will focus on spotted microarrays.

Spotted arrays. The term “spotted microarray” refers to a 2-dimensional series of elements on a surface, where each element

of the array is an aliquot of some nucleic acid that has been deposited in its specific location [as opposed to being synthesized at that location, base by base *in situ* (22)]. Spotted arrays are manufactured using x-y-z robots that use hollow pins to deposit cDNAs (PCR products), clones, or short and long oligonucleotides onto specially coated glass microscope slides (18). Spot sizes range between 50 and 500 μm in diameter (depending on the type of pin used, e.g., pin and ring versus quill or slit pin) (23), and arrays that contain up to 80 000 spots can be obtained, however, lower density arrays are more common (from 100 elements per cm^2 to as many as 25 000 elements per cm^2). Gene sequences to be arrayed are selected from several public databases, which contain resources to access well-characterized genes and expressed sequence tags (ESTs) representative of genes of unknown function. The clones chosen are amplified from appropriate cDNA libraries by PCR and purified before spotting on the solid support.

The strength of this technology is its flexibility: enabling relatively rapid iteration of experiments, where the number of array elements is in the range of hundreds to low thousands, and where 20 to 100 custom designed arrays can be fabricated for a particular experiment. In addition to their lower price and this flexibility in design, spotted arrays offer the ability to interrogate many surface-bound probes simultaneously with a very small amount of labeled target sample applied in a small volume (in contrast to the use of large filters as substrates for hybridization that require labeled probes applied in much larger volumes). Simultaneous expression analysis of two biological samples, such as untreated cells compared with treated cells or healthy tissue compared with cancer, also offers an enormous advantage for any pairwise analysis. Another strength of the technology is in enabling the study of genes that have not yet been sequenced (i.e., spotting ESTs offers the potential for the discovery of new genes and defining their role in disease). The major disadvantage of spotted arrays, which is due to the variability in spot quality from slide to slide, is that they provide information only on the relative gene expression between specific cells or tissue samples as opposed to direct quantification of RNA expression.

Affymetrix GeneChips™. A competing microarray technology that has garnered much attention is the GeneChip (Affymetrix, Santa Clara, CA, USA). High-density oligonucleotide GeneChips are produced by synthesizing tens of thousands of short oligonucleotides *in situ* onto glass wafers, one nucleotide at a time, using a combination of photolithography and light-directed solid-phase DNA synthesis (16,22). Generally, GeneChips are

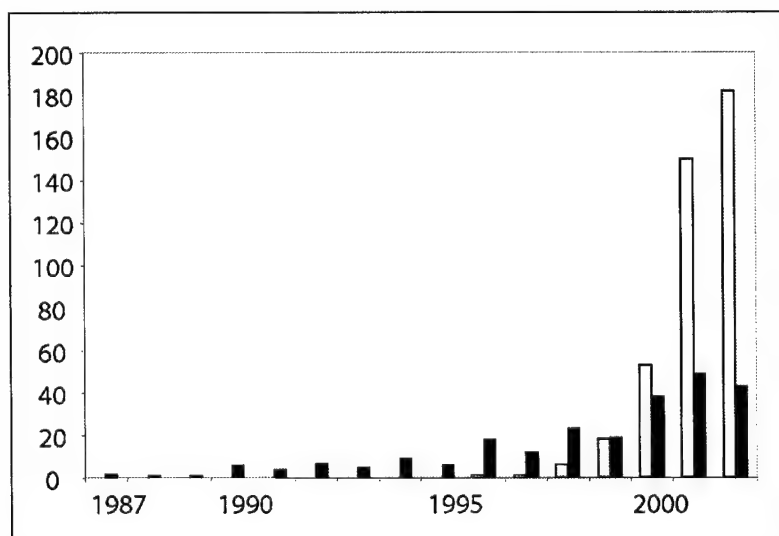


Figure 1. The growth in microarray publications in cancer research. The rapidly increasing number of publications describing DNA microarray analysis of cancers indicates the growing interest in this technique for cancer researchers (gray bars). Subtractive hybridization is another technique designed to isolate differentially regulated gene transcripts. By comparison with cancer microarray papers, the more labor-intensive subtractive hybridization method has generated fewer manuscripts over a much longer time period (black bars). Searches were performed at <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed> using the search strings “cancer and DNA and microarray” (gray bars) and “subtractive hybridization and cancer” (black bars). (Figure kindly supplied by Dr. Geraldine O’Neill of The Children’s Hospital Westmead, Westmead, NSW, Australia.)


designed with 16 to 20, preferably nonoverlapping 25-mers representing each gene on the array. Each oligonucleotide on the chip is matched with an almost identical one, differing only by a central, single-base mismatch. This mismatched oligonucleotide serves as an internal control for hybridization specificity and allows for determination of the degree of nonspecific binding by comparison of target binding intensity between the two partner oligonucleotides. GeneChip oligonucleotide probe arrays are ideally suited for applications involving analysis of large amounts of known sequence content (e.g., whole genome expression monitoring, where the vast majority of genes for a given organism have already been sequenced and can be synthesized as probes), where high quality is required for every array utilized, and where the same array design can be used repeatedly. Another advantage of Affymetrix GeneChips is their ability to measure the absolute expression of genes in cells or tissues. The sensitivity of the GeneChip enables the detection of mRNAs present at levels as low as 1 transcript in 100 000; at the extreme, sensitivity has been reported as low as 1:1 000 000 (16). Their disadvantages, in addition to their higher costs, include the inability to simultaneously compare, on the same array, the degree of expression of two related biological samples. In addition, oligonucleotide-based microarrays require a priori knowledge of the gene sequences and complex computational manipulation to convert the 40 feature signals into an actual expression value. More recently, oligonucleotide arrays have been developed that combine some of the flexibilities and qualitative advantages associated with the use of synthetic probe arrays with the benefits of simultaneous analysis afforded by spotted glass arrays (24).

Global Gene Expression in Tumor Biology

Identification of tumors by histology or immunohistochemistry yields information regarding tumor origin, tumor type, and tumor stage or grade, which to date still provide the most relevant prognostic information. However, staining patterns that are microscopically observed cannot predict the underlying biochemical and genetic events that are important for clinical outcome. Multiple genetic changes that result in the activation of oncogenes, increased production of growth factors, loss of growth inhibitory cytokines, or loss of function of tumor suppressor genes result in an imbalance of growth and survival regulation. The result of such derangements is a persistent pathologic communication state between the tumor cell and the host in which tumor cells successfully invade and metastasize because of somatic genetic progression that supports invasion and survival. While cancer researchers have made great strides in uncovering parallel molecular mechanisms that are involved between many different cancer types, the set of genes required for successful tumorigenesis is still debated. Clinically, it can be observed, for example, that two patients harboring similar tumors according to their cellular morphology may have two different clinical outcomes and/or may respond differently to the same treatment. Therefore, cancer researchers are applying cDNA and oligonucleotide microarrays to study several thousand genes at once to enhance our current tumor classification system (25–27), while concomitantly providing insights into pathogenesis, diagnosis, prognosis, therapeutic targets, and clinical outcome of tumors (16,26,28–31).

Tumor classification. In a landmark study, Golub and colleagues were among the first to correlate gene expression profiles with tumor classification (32). They were able to distinguish acute myeloid leukemia (AML) from acute lymphoblastic leukemia (ALL) based on gene expression alone in 36 out of 38 bone marrow samples. Yeoh and colleagues extended these types of studies and have demonstrated in pediatric ALL that expression profiling could not only accurately identify the known prognostically important leukemia subtypes, but could further enhance the ability to assess a patient's risk of failing therapy (33). Other groups have also used gene expression pattern analysis to classify, at the molecular level, ovarian tumors (34,35), breast tumors (36,37), B-cell lymphoma (38), cutaneous melanoma (28), and lung adenocarcinoma (39,40). Likewise, in a recent study analyzing molecular profiles of nonneoplastic and neoplastic prostate samples, Dhanasekaran and colleagues established a signature expression profile of healthy prostate, benign prostatic neoplasia, localized prostate cancer, and metastatic prostate cancer (41). These studies have demonstrated the feasibility of combining large-scale molecular analysis of expression profiles with classic morphologic and clinical methods of staging and grading cancer for better diagnosis and outcome prediction.

Tumor-specific molecular markers. As noted above, tumors can be divided clinically into those that originate in solid organs such as the kidney, bowel, or lung, and those that originate in the hematopoietic system. In fact, this distinction is helpful in considering how to apply array technology to expression monitoring of cancer, because issues of sample preparation and tissue heterogeneity are of particular concern in performing and interpreting experiments with solid tumors. Nevertheless, there is widespread optimism in the scientific community that microarray-based expression monitoring will help decipher the information hidden in sequenced genomes. There is a growing number of research groups that have focused on identifying subsets of genes that show differential expression between healthy tissues or cell lines and their tumor counterparts to identify biomarkers for several solid tumors, including ovarian cancer (42–48), oral cancer (49,50), endometrial cancer (51), melanoma (52), and colorectal cancer (53,54). Unfortunately, the realization from most of these studies is that the generation of data is not enough; one must be able to extract from it meaningful information about the system being studied. Furthermore, the issue of tissue homogeneity needs to be carefully considered. Global gene expression surveys assume that information of interacting cell populations can be globally investigated by analyzing gene expression patterns from heterogeneous tissue samples. Total RNA from tissue fragments (i.e., high tumor content versus low tumor content) is isolated, labeled, probed on an array, and the status of gene expression quantified. However, it is not hard to imagine that genes from cell subpopulations are also included and thus may contaminate the results. Therefore, much effort has been directed towards using microdissected cell populations, which in turn have their own advantages and disadvantages (55). Nevertheless, as more and more researchers rush to embrace the new techniques, one of the most important challenges facing these investigators will be determining which of the plethora of new differentially expressed genes is biologically relevant to the tumor system being studied. Even when rigorous efforts are made to minimize the number of variables in a mi-



croarray study, there may be an unmanageable number of differentially expressed genes that will contribute excessive background values. Therefore, the greatest contributions to understanding function will likely come from directly combining microarray data with other sources of genomic and biomedical information. Algorithms for integrating different types of data are already showing promise and integrating clinical data from patients' records has been proposed. By combining expression microarray analysis with other approaches, such as cytogenetic techniques, it may be possible to focus on a significantly smaller subset of genes of direct relevance to tumor biology (56). In practice, recent studies have combined expression arrays and comparative genomic hybridization array techniques on breast cancer cell lines and identified a limited number of genes that were both amplified and overexpressed (57,58). Efforts have also been initiated to apply in concert DNA microarrays and proteomics approaches to better understand the regulatory events involved in normal and disease processes (38). This is important given that the measurement of transcribed mRNA alone is not sufficient for the characterization of biological systems. It is clear that continued advancements in the comprehensive analysis of protein products in conjunction with the already mature methods in global measurement of mRNA expression will eventually help to uncover the molecular basis of cancer.

METHODS FOR ANALYZING MICROARRAY DATA

The fundamental change typifying cancer is the clonal proliferation of a single cell, which requires activation of the cell cycle and the overriding of cellular programs (e.g., apoptosis) that guard against proliferation. A number of specific cellular malfunctions have been identified, including loss of function of key tumor suppressor genes such as p53 (59-61), gain of function mutations of cell growth regulators such as c-KIT (62,63), and activation of proto-oncogenes such as c-MYC (64). Since cells within complex organisms have built-in safeguards against proliferation, a single adverse event in the cell is generally not adequate to initiate cancer (65), but two or more events are required to take a cell from normal to malignant, often including a definite sequence of cellular events (66). During these changes the expression levels of numerous genes are expected to undergo changes as the cell adjusts itself to a new program. These changes take place within a background of numerous routine expression changes or even stochastically varying levels of gene expression. The issue for microarray measurements then becomes identifying changes related to the process of interest, e.g., tumorigenesis, in a background of routine differences in expression between individual sets of cells. Such sets of cells may be measurements from different patients, different times in the

Review

same individual, or different tissue types within an individual.

The multitude of possible cellular events that can give rise to cancer lead to a diversity in types of cancer not seen in other diseases. Often cancers that appear histologically similar can have dramatically different responses to standard therapies and different courses of development (67). Since these differences in behavior are likely to be reflected in differences in the set of genes expressed, one promising use for microarrays is to more finely differentiate cancers using gene expression levels to bolster standard histology. The goal is to allow therapies, including the level of aggressiveness of treatment, to more closely match the underlying disease, hopefully reducing side effects in cases of low risk and increasing cure rates in cases of high risk. In this case the goal is identification of an expression profile that can be used to type the cancer.

Methods of Analysis for Classification

To use microarray measurements to distinguish cancer types, it is necessary to identify patterns for each subtype that uniquely distinguish that subtype from all others. This requires some form of pattern recognition, with the methods used generally borrowed from the fields of statistics, computer science, and data mining. Since the data generated in a microarray experiment usually comprise a matrix of measurements (Figure 2), methods of analysis developed for matrix manipulations can be applied to microarray measurements as well. Generally the individual rows of the matrix identify different genes or nucleotide sequences with the individual columns identifying different conditions. For the purposes of cancer identification, the columns would represent different tumors, patients, or tissues. The goal for an analysis aimed at identifying tumor profiles would be to divide the conditions into sets linked to tumor or tissue type based on expression levels of the genes. Two general types of methods can be used. The first type is unsupervised learning, in which there is no a priori identification of the group in which a condition (e.g., sample) should be placed (i.e., the type of tumor). The second type is supervised learning in which a well-defined learning set exists (e.g., each sample identified as to type of tumor) to allow an algorithm to develop a test that will be applied to additional samples of unknown type.

Unsupervised learning includes a set of methods that attempt to define groups within the data, often called clusters. These clusters are then interpreted in light of other information. For instance, in a study of gene expression in the NCI60 cell lines, the clusters identified could be linked to the tissue of origin of the tumor (68). The algorithms for clustering include a number of different methods, such as hierarchical and K-means clustering (69–71), bootstrapping or jackknife methods (72,73), methods which test for poor data (74), clustering based on statistical methods such as singular value decomposition or principal component analysis (75), fuzzy clustering methods (76,77), and self-organizing maps (78). Each of these methods has its own advantages and limitations. While there has been success in the use of these algorithms, no clear theoretical advantage exists for any given algorithm in application to expression data, so the safest approach is to apply multiple algorithms and to try to determine which functions best on the specific data set.

Supervised methods are generally borrowed from the field of

computer science and machine learning. These methods require a training set where the identification of each condition has been made by other means. For example, in their analysis, J. Khan et al. first trained the algorithm using a subset of the full data where the identification of the malignancy had already been made (79). The supervised learning methods that have been applied to gene expression data include support vector machines (SVM) (80), multidimensional scaling (MDS) (26), and artificial neural networks (79). The main goal behind these algorithms is to define a set of criteria that allow the maximal separation of the conditions into the correct types.

Specifics of Algorithms for Unsupervised Learning

Because of its ready availability due to the release of an easily used version by Eisen and colleagues (69), the pattern recognition algorithm most widely applied to microarray data is hierarchical clustering. Hierarchical clustering measures similarity between the genes based on expression profiles across conditions. This measurement requires a definition of similarity that is provided by a metric. The most common metric is Euclidean distance, which measures how different two profiles are taking into account the magnitude of signals (i.e., the vector [10, 20, 30] is very distant from the vector [1, 2, 3]). Another common distance metric is Pearson correlation coefficient, which measures the shape or direction (i.e., the vector [10, 20, 30] is at the same point as the vector [1, 2, 3]). The advantage of Pearson correlation as a metric is that genes that vary together but with one

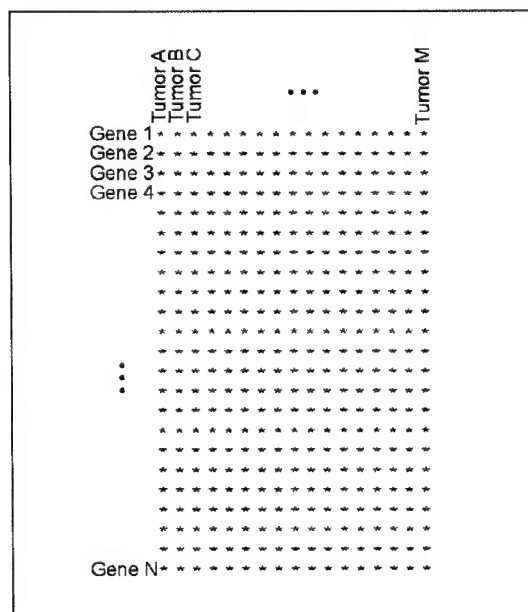


Figure 2. Typical form for microarray data in a cancer-related study following normalization and combining of replicates. The normalized data is a set of values of the expression level for each gene in a number of different conditions, typically in a number of tumors or tissue types.

REVIEW

much more highly expressed will still cluster together. However, this metric can miss coordinated changes if the changes are comparable in each gene but their background levels of expression vary by different amounts. In addition to the metric, there are multiple methods of building the clusters that vary on use of centroids, nearest members, or farthest members when measuring distances between and among clusters and expression profiles. The details of these issues are beyond the scope of this work, but can be explored in the literature (81,82). Finally, hierarchical clustering generally does not provide clusters but instead provides measures of distance between multiple possible groupings of genes. Generally this is shown as a dendrogram and it is left to the user to determine at what level to cut the tree to provide clusters (as shown in Figure 3). This level of control provides power but also makes the method prone to artifactual errors, so that some statistical analysis might be required to determine that the clusters are meaningful.

A second common clustering algorithm borrowed for use with microarray data is K-means clustering (83,84). With this method, the number of clusters must be chosen prior to application of the algorithm. The algorithm then identifies the cluster distribution and assigns members in a way to minimize overall the distances of each point from its cluster. Again a distance metric and method for identifying how to measure the distance of a point relative to the cluster must be chosen.

One problem faced by these clustering methods is the possibility that a single shared bad data point can lead to clustering

together two truly unrelated expression patterns. To address this issue, Heyer and colleagues developed a clustering technique that uses jackknife correlation as part of the distance metric (74). In this way, individual data points cannot dominate the underlying clustering.

A further problem with clustering is the issue of the correct number of clusters. In an attempt to answer this problem, Lukashin and Fuchs introduced a method that uses simulated annealing to estimate the number of clusters (71). Simulated annealing is a method borrowed from statistical mechanics that models the slow cooling of a physical system (85), effectively moving from a region where the data plays a minor role in determining the structures of the model to one where the data dominates.

Once the clusters are chosen, there remains the issue of statistical significance. Clearly the wide variety of parameters available in most clustering methods raises the issue of identifying clusters by chance. To validate the results of clustering, Kerr and Churchill introduced a method to use bootstrapping analysis to interpret the results of clustering (72). Bootstrapping estimates the data and noise, which is done here through an analysis of variance (ANOVA) model (86), and clusters models of the data repeatedly. In this way the reliability of the identified clusters can be estimated.

As noted previously, microarray data often take the form of a matrix of measurements, allowing the application of numerous techniques developed for matrix analysis. One such method, sin-

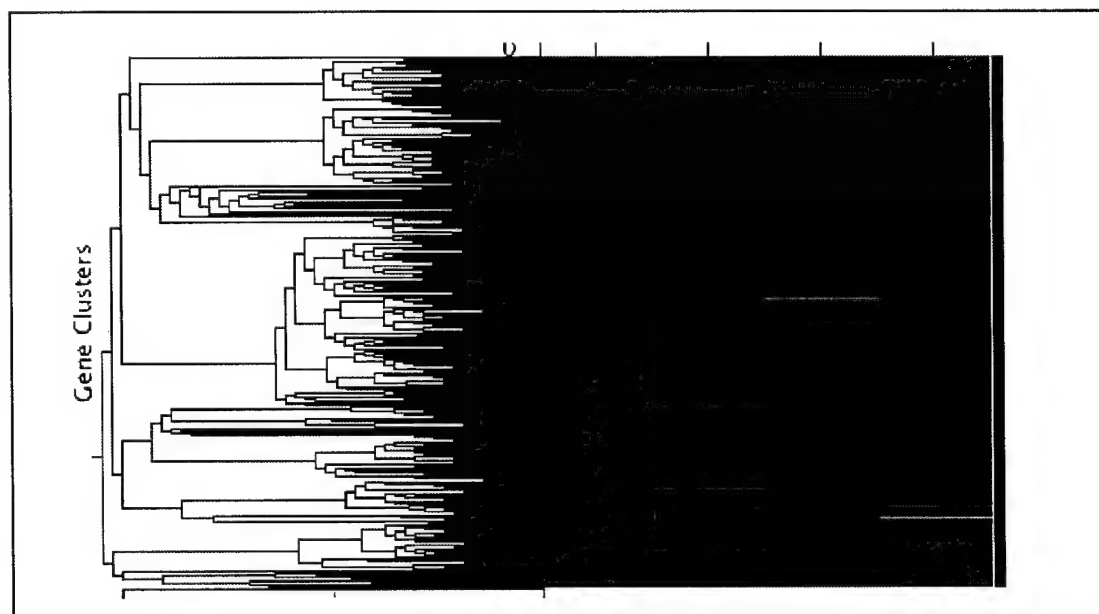


Figure 3. Hierarchical clustering of genes across four conditions. The most widely used pattern recognition method in microarray analysis is hierarchical clustering, in which genes are compared by expression profile across multiple conditions. In this image generated by BioDiscovery's GeneSight™ (Marina Del Rey, CA, USA), green represents lowered expression relative to the control and red represents higher expression. The dendrogram on the left indicates the magnitude of the difference between the expression profiles according to the metric. Clustering requires deciding where to cut the dendrogram to define clusters, which can therefore be a few clusters with many genes (cut on far left) or many clusters with few genes (cut on far right).

gular value decomposition (SVD, or principal component analysis) has been used to create a clustering method as well (75). In SVD, the matrix of measurements is converted into a matrix of principal components and a matrix of scores. The first principal component is usually chosen to maximize the amount of variance within the data explained by that single component, with all other components chosen orthogonal to all preceding components and in a way to maximize the amount of the remaining variance explained. In this way the data can be reduced in dimensionality, however the number of components to retain is still problematic. The scores then represent how to reconstruct the expression profile of each gene from the principal components, so that plotting the scores for the first three principal components can lead to clustering of the genes into similar groups in three dimensions.

Self-organizing maps are an additional method for clustering (78), which allows the user to guide the analysis by laying out a grid of nodes. These nodes are then moved to match the distribution of the data and each individual data point (i.e., each gene expression profile) is assigned to a node.

All of the methods looked at so far place each gene into a single cluster. This is similar to the logical operation of true (in cluster) or false (not in cluster). However, fuzzy logic systems have been developed that allow intermediate logical assignments between true and false. These methods lead to a clustering method, fuzzy clustering (76,77), that allows genes to be placed into multiple groups with an assignment to the group between 0 (not in group) and 1 (completely within group). This method has the advantage of allowing each gene the possibility of belonging to multiple clusters, which may reflect the underlying biology better (87).

Specifics of Algorithms for Supervised Learning

One of the most widely used machine learning algorithms is

an artificial neural network. These methods were adopted early in the development of microarray analysis in cancer research to identify classification criteria for tumors (79). These algorithms are discussed in this issue by Ringnér and Peterson. Essentially the method involves setting up a network of nodes in layers, with each node fed by all nodes in the previous layer. The strength of the influence of one node on another is tuned by the algorithm to give a final output most matching the predetermined assignment of each sample to a group. Ideally, this results in a nonlinear system tuned to identify the class of a new unknown sample. One drawback to these systems is that it can be hard to interpret the results biologically because of the multiple layers and nonlinearities involved in the neural network.

MDS is a method of mapping complex data into lower dimensional space in a way that maximizes the separation between classes. The classes are predefined (e.g., type of cancer) and the result is a mapping that can be applied to new data to identify the class to which an unknown sample belongs (26).

SVMs are another method of identifying features that separate two predefined classes under study (80). The features are formed by looking at the relationships between each two points in the multidimensional data set. The hyperplane that best separates those feature points is determined and used to classify new samples.

One problem with these methods is that they are trained on a specific set of data measured using a single platform (often within a single laboratory) and including a specific set of genes. For diagnostics to be useful, they must allow for measurements to take place at many locations and ideally allow different platforms for measurement (RT-PCR, Affymetrix GeneChip, Rosetta oligo arrays, etc.). One attempt to allow such broad measurement methods is to create classifiers with the validity of the classification tied to the underlying biology regardless of the measurement method (88). Such a method is described in this issue by Scharpf and colleagues. Essentially to overcome inherent differences in the quantitative reliability of different platforms, a binary classifier is used that measures the probabilities of expression or nonexpression of each gene in each sample based on the data. These are then used in a supervised learning method to identify a subset of genes that together by their presence or absence identify the cancer type.

Methods of Analysis for Exploring Signaling Pathways

As noted previously, cancer often arises from specific malfunctions in genes involved in cellular signaling. This leads to undesired changes in the activity of specific signaling pathways. Since these pathways generally control gene transcription, these changes are expected to lead to changes in gene expression. The methods discussed so far aim to identify signatures of cancer and type, but do not provide a method of interpreting these patterns in light of signaling pathways. The reason for this is shown diagrammatically in Figure 4.

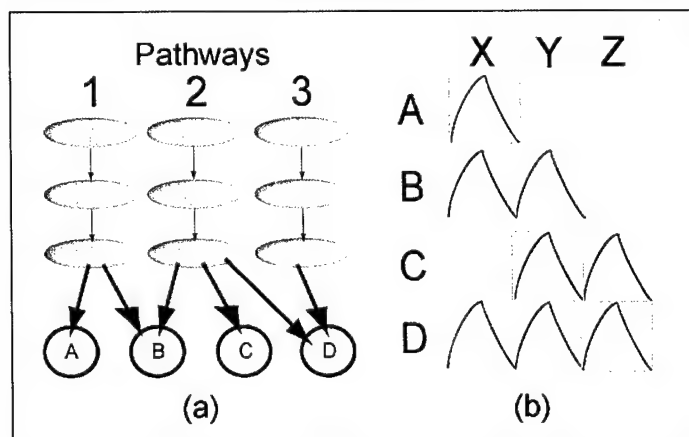


Figure 4. The link from signaling pathways to expression response. This highly simplified diagram demonstrates the problem in using expression analysis for identification of pathway activation even in simple cases (here all cross-talk is pushed to the point of transcription for clarity). The pathways in (a) are transiently active in the order 1&3 (X), 2 (Y), 3 (Z), resulting in the transcriptional profiles in (b). Algorithms that are unable to assign genes in B and C to multiple groups cannot recover the link to the underlying processes or even the correct number of processes.

Review

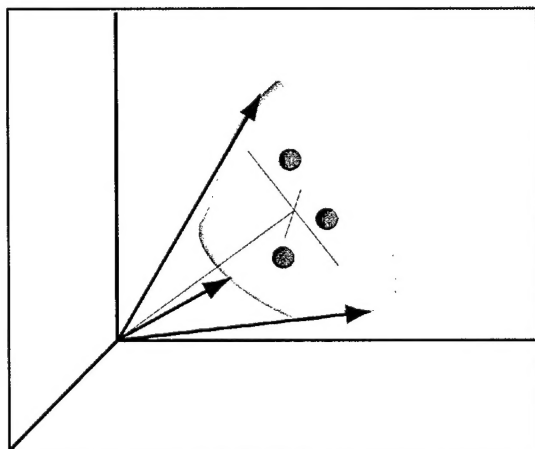


Figure 5. Comparison of methods of identifying patterns in multidimensional gene expression data. The blue lines represent the PCA solution; the green disks represent fuzzy clusters; and the red vectors represent the desired, physiologically significant patterns spanning the multidimensional data.

Since one result of the activation of a signaling pathway is transcription of a set of genes that generally overlap substantially with genes activated by other pathways, the clustering techniques cannot identify the set related to a specific pathway. This makes it difficult to link expression changes to signaling pathways.

However, pathway information is critical not only to the understanding of cancer development, but also to the design of effective therapeutics in the treatment of cancer. Present cancer treatments, such as radiotherapy and chemotherapy, result in substantial collateral damage to healthy tissues. Targeted therapies would try to alter behavior in a cell-specific manner, affecting only cancer cells. The creation of these therapies will require a detailed understanding of how disrupting specific cellular pathways affects downstream events in cells and an understanding of the signaling and metabolic networks to avoid unintended side effects in treatment (e.g., disrupting a pathway in a healthy cell leading to damage to healthy tissues).

One method of attacking this problem is to model the known pathways and analyze the expression data in light of the existing knowledge of the signaling pathways. The models can be quite detailed, including such information as rate constants, transcriptional initiation and elongation, etc. (89) or highly simplified, merely tracing connections (90). With the provided net-

work model, the expression data can be used to validate or choose between alternative models (91). The problem that arises with cancer is that the necessary network models remain poorly defined. While there has been some success with limited models (92,93), the gaps in knowledge of signaling networks and potential interactions appear too large to allow widespread use of these models presently (94).

If detailed network models are unavailable, the first step is instead to decompose the observed transcriptional responses into overlapping sets of genes. As can be seen in Figure 4, the desired form for the transcriptional response is generally a minimal structure (as shown in the boxes), since the desired decomposition is into minimal units related to a single pathway or, at least, minimal units related to all synchronized pathways within the data. Mathematically this is equivalent to identifying the minimal edge vectors bounding the data in a multidimensional space (as depicted in Figure 5). Methods that have been proposed to handle the multiple coregulation problem, such as principal component analysis (PCA) and fuzzy clustering (FC), generally operate in the center of the distribution. In addition, the requirement that all principal components (PCs) be orthogonal limits the usefulness of PCA in this type of analysis (see blue lines in Figure 5), since the basis vectors are typically nonorthogonal. For FC, the problem is that the bulk of the data will generally lie on the interior of the distribution rather than the edges, so that the seed clusters will not be located correctly for recovery of the fundamental basis vectors (see green disks in Figure 5).

One approach to recover the nonorthogonal basis vectors (i.e., minimal expression patterns) directly is to use Markov chain Monte Carlo to search through all probable solutions for the one that best reproduces the data. Bayesian Decomposition (87) uses a Bayesian statistical framework (95) that includes a goal of minimizing the structure in the model to identify the minimal patterns and to simultaneously determine the distribution of genes within them. Working on public domain data, it has successfully identified cell cycle regulated genes in the Stanford yeast cell cycle data set (96,97), including multiple coregulation of a number of kinases (87), isolated the transcriptional mating response pattern in *Saccharomyces cerevisiae* independent of any knowledge of the proteins involved in the signaling pathway (98) in the Rosetta deletion mutant data set (99), and identified patterns and specific genes expressed in mouse organs (100) from the Project Normal data (101). This method is limited since it is driven by the data without a model of the network. As such, it can only recover information with the resolution of the input data—that is, if during the creation of the data three signaling pathways were equally active at all times, those three pathways cannot be identified as separate processes.

CONCLUSION

In human cancer, the use of microarrays is starting to provide major insights into the genesis, progression, and response to therapy. With the availability of the human genome sequence, the goal of examining the entire transcriptome of normal and malignant cells is now possible. In combination with the continued development of the informatics and data analysis tools necessary to interrogate and interpret these data, this approach

promises to revolutionize how we think about cancer and how we are likely to detect and treat cancer in the future. By ultimately combining genomic with proteomic approaches, the information gained from these integrative kinds of studies may lead to the identification of secreted proteins that can be used for sensitive, presymptomatic molecular diagnosis. It may also be possible to find membrane-bound proteins with functions ideally suited for protein therapeutic approaches. In this regard, understanding tumor subclassification is paramount, because expecting cell surface markers that are common to all tumors within a specific malignancy is probably naïve. However, if two or three subclasses of that malignancy can be discerned, it is likely that the strongest predictor genes will qualify as candidates for use in molecular diagnosis. Although there are a variety of technology options available from which to choose, it is clear that the impact of microarray technology on genomic research will be enormous. Going forward, one or another form of this technology will likely be used in all laboratories interested in the study of complex genetic information.

REFERENCES

- Greenlee, R.T., T. Murray, S. Bolden, and P.A. Wingo. 2000. Cancer Statistics, 2000. *CA Cancer J. Clin.* 50:7-33.
- Jemal, A., T. Murray, A. Samuels, A. Ghafoor, E. Ward, and M.J. Thun. 2003. Cancer Statistics, 2003. *CA Cancer J. Clin.* 53:5-26.
- Simmonds, M.A. 2003. Cancer Statistics, 2003: Further decrease in mortality rate, increase in persons living with cancer. *CA Cancer J. Clin.* 53:4.
- Lander, E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* 409:860-921.
- Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, et al. 2001. The sequence of the human genome. *Science* 291:1304-1351.
- Aebersold, R., L.E. Hood, and J.D. Watts. 2000. Equipping scientists for the new biology. *Nat. Biotechnol.* 18:359.
- Liang, P. and A.B. Pardee. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967-971.
- Martin, K.J. and A.B. Pardee. 1999. Principles of differential display. *Methods Enzymol.* 303:234-258.
- Lisitsyn, N. and M. Wigler. 1993. Cloning the differences between two complex genomes. *Science* 259:946-951.
- Diatchenko, L., Y.-F. Lau, A. Campbell, A. Chienchik, F. Moqadam, B. Huang, S. Lukyanov, K. Lukyanov, et al. 1996. Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* 93:6025-6030.
- Wang, X. and G.Z. Feuerstein. 2000. Suppression subtraction hybridization: application in the discovery of novel pharmacological targets. *Pharmacogenomics* 1:101-108.
- Velculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270:484-487.
- Agaton, C., P. Unneberg, M. Sievertzon, A. Holmberg, M. Ehn, M. Larsson, J. Odeberg, M. Uhlen, and J. Lundberg. 2002. Gene expression analysis by signature pyrosequencing. *Gene* 289:31-39.
- DeRisi, J.L., V.R. Iyer, and P.O. Brown. 1997. Exploring the metabolic and genetic control of gene expression on a genome scale. *Science* 278:680-686.
- Lashkari, D.A., J.L. DeRisi, J.H. McCusker, A.F. Namath, C. Gentile, S.Y. Hwang, P.O. Brown, and R.W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci. USA* 94:13057-13062.
- Lipshultz, R.J., S.P. Fodor, T.R. Gingeras, and D.J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21:20-24.
- Lockhart, D.J. and E.A. Winzler. 2000. Genomics, gene expression and DNA arrays. *Nature* 405:827-836.
- Schena, M., D. Shalon, R.W. Davis, and P.O. Brown. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA mi-



- croarray. *Science* 270:467-470.
19. Pollack, J.R., C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, and C.F. Williams. 1999. Genome-wide analysis of DNA copy number changes using cDNA microarrays. *Nat. Genet.* 23:41-46.
20. Frolov, A., A.H. Prowse, L. Vanderveer, B. Bove, H. Wu, and A.K. Godwin. 2002. DNA array-based method for detection of large rearrangements in the BRCA1 gene. *Genes Chromosomes Cancer* 35:232-241.
21. Haab, B.B. 2001. Advances in protein microarray technology for protein expression and interaction profiling. *Curr. Opin. Drug Discov. Dev.* 4:116-123.
22. Fodor, S.P., J.L. Read, M.C. Pirrung, L. Stryer, A.T. Lu, and D. Solas. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767-773.
23. Rose, S.D. 2002. Spotted arrays: technology overview, p. 3-14. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westboro, MA.
24. Okamoto, T., T. Suzuki, and N. Yamamoto. 2000. Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat. Biotechnol.* 18:438-441.
25. Afshari, C.A., E.F. Nuwaysir, and J.C. Barrett. 1999. Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res.* 59:4759-4760.
26. Khan, J., R. Simon, M. Bittner, Y. Chen, S.B. Leighton, T. Pohida, P.D. Smith, Y. Jiang, et al. 1998. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* 58:5009-5013.
27. Marx, J. 2000. Medicine. DNA arrays reveal cancer in its many forms. *Science* 289:1670-1672.
28. Bittner, M., P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, et al. 2000. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406:536-540.
29. DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, and J.M. Trent. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 1:4457-460.
30. Pease, A.C., D. Solas, E.J. Sullivan, M.T. Cronin, C.P. Holmes, and S.P. Fodor. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci. USA* 91:5022-5026.
31. Perou, C.M., S.S. Jeffrey, M. van de Rijn, C.A. Rees, M.B. Eisen, D.T. Ross, A. Pergamenschikov, C.F. Williams, et al. 1999. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* 96:9212-9217.
32. Golub, T.R., D.K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J.P. Mesirov, H. Coller, M.L. Loh, et al. 1999. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-537.
33. Yeoh, E.J., M.E. Ross, S.A. Shurtleff, W.K. Williams, D. Patel, R. Mahfouz, F.G. Behm, S.C. Raimondi, et al. 2002. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1:133-143.
34. Ono, K., T. Tanaka, T. Tsunoda, O. Kitahara, C. Kihara, A. Okamoto, K. Ohiai, T. Takagi, and Y. Nakamura. 2000. Identification by cDNA microarray of genes involved in ovarian carcinogenesis. *Cancer Res.* 60:5007-5011.
35. Schwartz, D.R., S.L. Kardia, K.A. Shedden, R. Kuick, G. Michailidis, J.M. Taylor, D.E. Misek, R. Wu, et al. 2002. Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Res.* 62:4722-4729.
36. Perou, C.M., T. Sorlie, M.B. Eisen, M. van de Rijn, S.S. Jeffrey, C.A. Rees, J.R. Pollack, D.T. Ross, et al. 2000. Molecular portraits of human breast tumours. *Nature* 406:747-752.
37. Sorlie, T., C.M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M.B. Eisen, et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* 98:10869-10874.
38. Celis, J.E., M. Kruhoffer, I. Gromova, C. Federiksen, M. Ostergaard, T. Thykjaer, P. Gromov, J. Yu, et al. 2000. Gene expression profiling: monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* 480:2-16.
39. Bhattacharjee, A., W.G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, et al. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* 98:13790-13795.
40. Garber, M.E., O.G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. van de Rijn, G.D. Rosen, et al. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* 98:13784-13789.
41. Dhanasekaran, S.M., T.R. Barrette, D. Ghosh, R. Shah, S. Varambally, K. Kurachi, K.J. Pienta, M.A. Rubin, and A.M. Chinnaiyan. 2001. Delineation of prognostic biomarkers in prostate cancer. *Nature* 412:822-826.
42. Ismail, R.S., R.L. Baldwin, J. Fang, D. Browning, B.Y. Karlan, J.C. Gasson, and D.D. Chang. 2000. Differential gene expression between normal and tumor-derived ovarian epithelial cells. *Cancer Res.* 60:6744-6749.
43. Mok, S.C., J. Chao, S. Skates, K. Wong, G.K. Yiu, M.G. Muto, R.S. Berkowitz, and D.W. Cramer. 2001. Prostatein, a potential serum marker for ovarian cancer: identification through microarray technology. *J. Natl. Cancer Inst.* 93:1458-1464.
44. Sawiris, G.P., C.A. Sherman-Baust, K.G. Becker, C. Cheadle, D. Teichberg, and P.J. Morin. 2002. Development of a highly specialized cDNA array for the study and diagnosis of epithelial ovarian cancer. *Cancer Res.* 62:2923-2928.
45. Shridhar, V., J. Lee, A. Pandita, S. Iturria, R. Avula, J. Staub, M. Morrissey, E. Calhoun, et al. 2001. Genetic analysis of early- versus late-stage ovarian tumors. *Cancer Res.* 61:5895-5904.
46. Shridhar, V., A. Sen, J. Chien, J. Staub, R. Avula, S. Kovats, J. Lee, J. Lillie, and D.I. Smith. 2002. Identification of underexpressed genes in early- and late-stage primary ovarian tumors by suppression subtraction hybridization. *Cancer Res.* 62:262-270.
47. Welsh, J.B., P.P. Zarrinkar, L.M. Sapinoso, S.G. Kern, C.A. Behling, B.J. Monk, D.J. Lockhart, R.A. Burger, and G.M. Hampton. 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci. USA* 98:1176-1181.
48. Wong, K.-K., R.S. Cheng, R.S. Berkowitz, and S.C. Mok. 2002. Gene expression analysis of ovarian cancer cells by cDNA microarrays, p. 127-138. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westborough, MA.
49. Alevizos, I., M. Mahadevappa, X. Zhang, H. Ohyama, Y. Kohno, M. Posner, G.T. Gallagher, M. Varvares, et al. 2001. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* 20:6196-6204.
50. Todd, R., J.S. Gutkind, E.J. Shillitoe, and D. Wong. 2002. Solid tumors: microarray analysis of oral cancers, p. 139-153. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westborough, MA.
51. Du, F., M. Mahadevappa, J.A. Warrington, and A.M. Bowcock. 2002. Gene expression changes in endometrial cancer, p. 113-125. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westborough, MA.
52. Clark, E.A., T.R. Golub, E.S. Lander, and R.O. Hynes. 2000. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* 406:532-535.
53. Hegde, P., R. Qi, R. Gaspard, K. Abernathy, S. Dharap, J. Earle-Hughes, C. Gay, N.U. Nwokekeh, et al. 2001. Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res.* 61:7792-7797.
54. Notterman, D.A., C.J. Shawber, and A.J. Levine. 2002. Tumor biology and microarray analysis of solid tumors: colorectal cancer as a model system, p. 81-111. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westborough, MA.
55. Pawletz, C.P. and L.A. Liotta. 2002. Tumor classification: gene analysis using DNA microarrays, p. 61-77. In J.A. Warrington, C.R. Todd, and D. Wong (Eds.), *Microarrays and Cancer Research*. Eaton Publishing, Westborough, MA.
56. Bayani, J., J.D. Brenton, P.F. Macgregor, B. Beheshti, M. Albert, D. Nalainathan, J. Karaskova, B. Rosen, et al. 2002. Parallel analysis of sporadic primary ovarian carcinomas by spectral karyotyping, comparative genomic hybridization, and expression microarrays. *Cancer Res.* 62:3466-3476.
57. Barlund, M., F. Forozan, J. Kononen, L. Bubendorf, Y. Chen, M.L. Bittner, J. Torhorst, P. Haas, et al. 2000. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.* 92:1252-1259.
58. Monni, O., M. Barlund, S. Mousses, J. Kononen, G. Sauter, M. Heiskanen, P. Paavola, K. Avela, et al. 2001. Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. USA* 98:5711-5716.
59. Carson, D.A. and A. Lois. 1995. Cancer progression and p53. *Lancet*

- 3461009-1011.
60. Fisher, D.E. 2001. The p53 tumor suppressor: critical regulator of life and death in cancer. *Apoptosis* 6:7-15.
 61. Harris, C.C. 1996. Structure and function of the p53 tumor suppressor gene: clues for rational cancer therapeutic strategies. *J. Natl. Cancer Inst.* 88:1442-1455.
 62. Hirota, S. 2001. Gastrointestinal stromal tumors: their origin and cause. *Int. J. Clin. Oncol.* 6:1-5.
 63. Hirota, S., K. Isozaki, Y. Moriyama, K. Hashimoto, T. Nishida, S. Ishiguro, K. Kawano, M. Hanada, et al. 1998. Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science* 279:577-580.
 64. Pelengaris, S., M. Khan, and G. Evan. 2002. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* 2:764-776.
 65. Knudson, A.G. 1971. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl. Acad. Sci. USA* 68:820-823.
 66. Grady, W.M. and S.D. Markowitz. 2002. Genetic and epigenetic alterations in colon cancer. *Annu. Rev. Genomics Hum. Genet.* 3:101-128.
 67. Alizadeh, A.A., M.B. Eisen, R.E. Davis, C. Ma, I.S. Lossos, A. Rosenwald, J.C. Boldrick, H. Sabet, et al. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403:503-511.
 68. Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, et al. 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24:227-235.
 69. Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95:14863-14868.
 70. Getz, G., E. Levine, and E. Domany. 2000. Coupled two-way clustering analysis of gene microarray data. *Proc. Natl. Acad. Sci. USA* 97:12079-12084.
 71. Lukashin, A.V. and R. Fuchs. 2001. Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters. *Bioinformatics* 17:405-414.
 72. Kerr, M.K. and G.A. Churchill. 2001. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* 98:8961-8965.
 73. Yeung, K.Y., D.R. Haynor, and W.L. Ruzzo. 2001. Validating clustering for gene expression data. *Bioinformatics* 17:309-318.
 74. Heyer, L.J., S. Kruglyak, and S. Yooseph. 1999. Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* 9:1106-1115.
 75. Alter, O., P.O. Brown, and D. Botstein. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* 97:10101-10106.
 76. Futschik, M.E. and N.K. Kasabov. 2002. Fuzzy clustering of gene expression data. p. 414-419. 2002 IEEE International Conference on Fuzzy Systems. IEEE, Honolulu, HI.
 77. Gasch, A.P. and M.B. Eisen. 2002. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* 3:RESEARCH0059.
 78. Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub. 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96:2907-2912.
 79. Khan, J., J.S. Wei, M. Ringner, L.H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, et al. 2001. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* 7:673-679.
 80. Brown, M.P., W.N. Grundy, D. Lin, N. Cristianini, C.W. Sugnet, T.S. Furey, M. Ares, Jr., and D. Haussler. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* 97:262-267.
 81. Jajuga, K., A. Sokolowski, and H.-H. Bock. 2002. Classification, Clustering, and Data Analysis. Springer-Verlag, New York.
 82. Kaufman, L. and P.J. Rousseeuw. 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.
 83. Soukas, A., P. Cohen, N.D. Succi, and J.M. Friedman. 2000. Leptin-specific patterns of gene expression in white adipose tissue. *Genes Dev.* 14:963-980.
 84. Soukas, A., N.D. Succi, B.D. Saatkamp, S. Novelli, and J.M. Friedman. 2001. Distinct transcriptional profiles of adipogenesis in vivo and in vitro. *J. Biol. Chem.* 276:34167-34174.
 85. Geman, S. and D. Geman. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6:721-741.
 86. Kerr, M.K., C.A. Afshari, L. Bennett, P. Bushel, J. Martinez, N.J. Walker, and G.A. Churchill. 2002. Statistical analysis of a gene expression microarray experiment with replication. *Statistica Sinica* 12:203-218.
 87. Moloshok, T.D., R.R. Klevecz, J.D. Grant, F.J. Manion, W.I. Speier, and M.F. Ochs. 2002. Application of Bayesian decomposition for analysing microarray data. *Bioinformatics* 18:566-575.
 88. Parmigiani, G., E. Garrett, R. Anbazhagan, and E. Gabrielson. 2002. A statistical framework for expression-based molecular classification in cancer. *J. R. Statistical Soc. [B]* 64:717-736.
 89. Gilman, A. and A.P. Arkin. 2002. Genetic "code": representations and dynamical models of genetic components and networks. *Annu. Rev. Genomics Hum. Genet.* 3:341-369.
 90. Kaufman, S. 1993. The Origins of Order: Self-Organization and Selection in Evolution. Oxford University Press, Oxford.
 91. Hartemink, A.J., D.K. Gifford, T.S. Jaakkola, and R.A. Young. 2001. Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.* 422-433.
 92. Tanaka, H., H. Arakawa, T. Yamaguchi, K. Shiraiishi, S. Fukuda, K. Matsui, Y. Takei, and Y. Nakamura. 2000. A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* 404:42-49.
 93. Yu, J., L. Zhang, P.M. Hwang, C. Rago, K.W. Kinzler, and B. Vogelstein. 1999. Identification and classification of p53-regulated genes. *Proc. Natl. Acad. Sci. USA* 96:14517-14522.
 94. Harkin, D.P. 2000. Uncovering functionally relevant signaling pathways using microarray-based expression profiling. *Oncologist* 5:501-507.
 95. Sibisi, S. and J. Skilling. 1997. Prior distributions on measure space. *J. Royal Statistical Soc. [B]* 59:217-235.
 96. Cho, R.J., M.J. Campbell, E.A. Winzler, L. Steinmetz, A. Conway, L. Wodicka, T.G. Wolfsberg, A.E. Gabrielson, et al. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2:65-73.
 97. Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9:3273-3297.
 98. Bidaut, G., T.D. Moloshok, J.D. Grant, F.J. Manion, and M.F. Ochs. 2002. Bayesian Decomposition analysis of gene expression in yeast deletion mutants, p. 105-122. In K. Johnson and S. Lin (Eds.), *Methods of Microarray Data Analysis II*. Kluwer Academic, Boston.
 99. Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, et al. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102:109-126.
 100. Moloshok, T.D., D. Datta, A.V. Kossenkova, and M.F. Ochs. Bayesian decomposition classification of the project normal data set. In K. Johnson and S. Lin (Eds.), *Methods of Microarray Data Analysis III*. Kluwer Academic, Boston. (In press.)
 101. Pritchard, C.C., L. Hsu, J. Delrow, and P.S. Nelson. 2001. Project normal: defining normal variance in mouse gene expression. *Proc. Natl. Acad. Sci. USA* 98:13266-13271.

Address correspondence to:

Dr. Michael F. Ochs
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111, USA
e-mail: m_ochs@fccc.edu
or
Dr. Andrew K. Godwin
Fox Chase Cancer Center
7701 Burholme Avenue
Philadelphia, PA 19111, USA
e-mail: a_godwin@fccc.edu